



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA PODNIKATELSKÁ

FACULTY OF BUSINESS AND MANAGEMENT

ÚSTAV INFORMATIKY

INSTITUTE OF INFORMATICS

**VYTVOŘENÍ MONITOROVACÍHO A PROFILOVACÍHO
ŘEŠENÍ NAD BI SYSTÉMEM**

MONITORING AND PROFILING SOLUTION FOR THE BI SYSTEM

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Matej Veselovský

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Jiří Kříž, Ph.D.

BRNO 2017

Zadání diplomové práce

Ústav: Ústav informatiky
Student: **Bc. Matej Veselovský**
Studijní program: Systémové inženýrství a informatika
Studijní obor: Informační management
Vedoucí práce: **Ing. Jiří Kříž, Ph.D.**
Akademický rok: 2016/17

Ředitel ústavu Vám v souladu se zákonem č. 111/1998 Sb., o vysokých školách ve znění pozdějších předpisů a se Studijním a zkušebním řádem VUT v Brně zadává diplomovou práci s názvem:

Vytvoření monitorovacího a profilovacího řešení nad BI systémem

Charakteristika problematiky úkolu:

Úvod
Cíle práce, metody a postupy zpracování
Teoretická východiska práce
Analýza současného stavu
Vlastní návrhy řešení
Závěr
Seznam použité literatury
Přílohy

Cíle, kterých má být dosaženo:

Cílem práce je tvorba monitorovacího řešení nad stávajícím BI systémem.

Základní literární prameny:

KIMBALL, Ralph a Margy ROSS. The data warehouse toolkit: the definitive guide to dimensional modeling. Third edition. ISBN 978-1-118-53080-1.

KNIGHT, Brian. Professional SQL server 2008 integration services. Indianapolis, IN: Wiley Pub., c2008. ISBN 978-047-0247-952.

NOVOTNÝ, Ota, Jan POUR a David SLÁNSKÝ. Business intelligence: jak využít bohatství ve vašich datech. Praha: Grada, 2005. Management v informační společnosti. ISBN 80-247-1094-3.

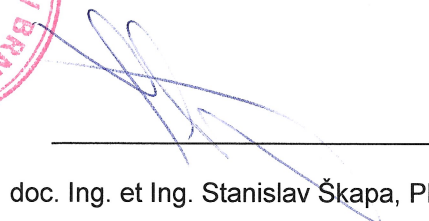
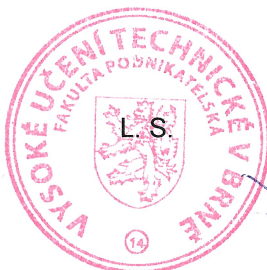
SIVAKUMAR HARINATH .. [ET AL.]. Professional Microsoft SQL Server Analysis Services 2008 with MDX. Indianapolis: Wiley, 2009. ISBN 978-047-0247-983.

Termín odevzdání diplomové práce je stanoven časovým plánem akademického roku 2016/17.

V Brně, dne 28. 2. 2017



doc. RNDr. Bedřich Půža, CSc.
ředitel



doc. Ing. et Ing. Stanislav Škapa, Ph.D.
děkan

Abstrakt

Táto diplomová práca sa zaoberá monitorovaním a profilovaním multidimenzionálnej databázy a reportov. Za týmto účelom bol vytvorený projekt v Microsoft SQL Server Integration Services, ktorý obsahuje 4 ETL balíčky. Práca je rozdelená na tri hlavné časti. Prvou je teoretická časť, kde sú vysvetlené pojmy a znalosti potrebné k úspešnému splneniu cieľa. V druhej časti sa nachádza analýza firmy, pre ktorú je dané riešenie vyvíjané a v tretej časti je popísaný samotný návrh a jeho riešenie. Praktická časť bola spracovaná v programoch MS Visual Studio 2015, MS SQL Management Studio 2012 a Power BI.

Abstract

This master's thesis focuses on monitoring and profiling multidimensional database and reports. The project was created in MS SSIS to this purpose and it contains 4 ETL packages. The thesis is divided into three main sections. First section consists of theoretical background needed to accomplish the goal. In second section there is analysis of the company for which is the solution created and in the third part of the thesis there is described the proposal and it's solution. Proposal of the solution and the solution itself was created in MS Visual Studio 2015, MS SQL Management Studio and Power BI.

Klíčové slová

Business Intelligence, Dátový sklad, SSIS, SSAS, SSRS, datová kocka, Power BI, Excel, SQL, MDX

Key words

Business Intelligence, Data warehouse, SSIS, SSAS, SSRS, data cube, Power BI, Excel, SQL, MDX

Bibliografická citácia

VESELOVSKÝ, M. *Vytvoření monitorovacího a profilovacího řešení nad BI systémem*. Brno: Vysoké učení technické v Brně, Fakulta podnikatelská, 2017. 75 s.
Vedoucí diplomové práce Ing. Jiří Kříž, Ph.D..

Čestné prehlásenie

Prehlasujem, že predložená diplomová práca je pôvodná a spracoval som ju samostatne. Prehlasujem, že citácie použitých prameňov sú úplné, a že som vo svojej práci neporušil autorské práva (v zmysle Zákona č. 121/2000 Sb., o práve autorskom a o právach súvisiacich s právom autorským).

V Brne dňa 10. Mája 2017

.....

Veselovský Matej

Pod'akovanie

Rád by som pod'akoval spoločnosti Intelligent Technologies, hlavne Lubošovi Bednárovi a Radimovi Hampelovi, za to, že som pri nich mohol stráviť posledný rok štúdia a samozrejme i mojim kolegom, že mi za ten čas pomohli ako s prácou tak aj so štúdiom. Taktiež by som chcel pod'akovať svojej priateľke, že pri mne stála a bola mi oporou v týchto nie zrovna najľahších časoch.

Obsah

| | |
|-------------------------------------|----|
| Abstrakt..... | 4 |
| Abstract..... | 4 |
| Kľúčové slová..... | 5 |
| Key words | 5 |
| Bibliografická citácia..... | 6 |
| Čestné prehlásenie | 7 |
| PodĎakovanie | 8 |
| Obsah | 9 |
| Úvod..... | 14 |
| Popis spoločnosti | 15 |
| 1 Teoretické východiská práce | 16 |
| 1.1 MS SQL Server | 16 |
| 1.2 MS SQL Database engine | 16 |
| 1.3 MS SQL SRS..... | 17 |

| | |
|--|----|
| 1.4 MS SQL SSAS | 19 |
| 1.5 MS SQL SSIS..... | 19 |
| 1.6 Business Intelligence | 21 |
| 1.7 ETL..... | 21 |
| 1.8 Data flow task..... | 23 |
| 1.9 Tabuľky dimenzií | 23 |
| 1.10 Tabuľky faktov | 24 |
| 1.11 OLAP..... | 24 |
| 1.12 Jazyk SQL | 25 |
| 1.13 Jazyk MDX..... | 25 |
| 1.14 Multidimenzionálna databáza..... | 26 |
| 1.15 Data warehouse | 27 |
| Analýza súčasného stavu | 28 |
| 1.16 SLEPT analýza – analýza obecného okolia | 28 |
| 1.16.1 Sociálne faktory | 28 |
| 1.16.2 Legislatívne faktory | 28 |
| 1.16.3 Ekonomické faktory..... | 28 |

| | |
|--|----|
| 1.16.4 Politické faktory..... | 29 |
| 1.16.5 Technologické faktory | 29 |
| 1.17 Porterov model piatich konkurenčných síl – analýza oborového okolia. 29 | |
| 1.17.1 Konkurencia prostredia..... | 29 |
| 1.17.2 Hrozba substitútov | 30 |
| 1.17.3 Zákazníci..... | 30 |
| 1.17.4 Dodávateľia | 30 |
| 1.17.5 Riziko vstupu nových konkurentov | 30 |
| 1.18 Model 7S – analýza interných faktorov | 30 |
| 1.18.1 Stratégia | 30 |
| 1.18.2 Štruktúra | 31 |
| 1.18.3 Informačné systémy..... | 31 |
| 1.18.4 Štýl riadenia | 31 |
| 1.18.5 Spolupracovníci | 31 |
| 1.18.6 Zdieľané hodnoty..... | 31 |
| 1.18.7 Schopnosti | 32 |
| 1.19 SWOT analýza | 32 |

| | |
|---------------------------------------|----|
| 1.19.1 Silné stránky | 32 |
| 1.19.2 Slabé stránky..... | 34 |
| 1.19.3 Príležitosti..... | 35 |
| 1.19.4 Hrozby | 35 |
| 1.20 Záver analýzy | 37 |
| 2 Vlastný návrh riešenia | 38 |
| 2.1 Popis navrhovaného riešenia | 38 |
| 2.2 Lewinov model..... | 38 |
| 2.2.1 Dôvod zmeny..... | 38 |
| 2.2.2 Agent zmeny | 39 |
| 2.2.3 Intervenčné oblasti..... | 39 |
| 2.2.4 Realizácia zmeny | 39 |
| 2.2.5 Zhodnotenie zmeny | 39 |
| 2.3 Zber dát..... | 39 |
| 2.3.1 OLAP Query Log | 40 |
| 2.3.2 Load Report Log..... | 41 |
| 2.3.3 Load Query Log..... | 44 |

| | |
|---|----|
| 2.3.4 Search for users..... | 45 |
| 2.3.5 Vytvorenie tabuliek faktov a dimenzií | 47 |
| 2.3.6 Vytvorenie multidimenzionálnej databázy | 53 |
| 2.3.7 Reporty | 55 |
| 2.3.8 Zobrazenie prehliadania reportov | 56 |
| 2.3.9 Zobrazenie vyt'aženosti kocky..... | 63 |
| 2.3.10 Reporty v MS Excel..... | 67 |
| 2.3.11 Ďalšie spracovanie | 68 |
| Záver | 69 |
| Zoznam použitých zdrojov | 70 |
| Zoznam obrázkov | 72 |
| Poznámky..... | 75 |

Úvod

Pri vytváraní produktov, či už to materiálnych alebo iných, pre svojich zákazníkov, nikdy nevieme ako s nimi naložia. Preto vznikla požiadavka zistiť čo a ako naši zákazníci používajú a tým sa tak obohatiť o vedomosti, ktoré sa dajú využiť pri budúcom vývoji alebo aj pri správe už vytvorených produktov.

Cieľom tejto práce je vytvorenie monitorovacieho riešenia nad OLAP kockou a reportami nasadenými na portáli, pomocou zberu dát o využívaní jednotlivých častí a ich zápisom do novovytvorenej štruktúry. Výstupom takto navrhnutého úložiska budú administratívne reporty, ktoré by mali pomôcť pri zobrazovaní takto nazbieraných dát. Ide hlavne o navrhnutie a vytvorenie prvotného riešenia, ktoré bude fungovať a slúžiť ako podklad pre ďalší vývoj resp. pre ďalšiu jeho úpravu. Hlavným cieľom tejto práce a samotného riešenia je zistenie a pochopenie zákazníka tak aby bolo možné ponúknuť mu čo najlepšie riešenie, ktoré bude mať pre neho maximálny prínos.

Popis spoločnosti

Firma Intelligent Technologies s.r.o sa zaoberá vývojom BI riešení pre svojich zákazníkov. To zahŕňa vývoj dátových skladov, analytických nástrojov, komplexných reportingových systémov, manažérskych informačných systémov a mnoho ďalších vecí. Spoločnosť Intelligent Technologies je partnerom spoločnosti Microsoft a drží množstvo ocenení a certifikátov na poli Business Intelligence systémov.



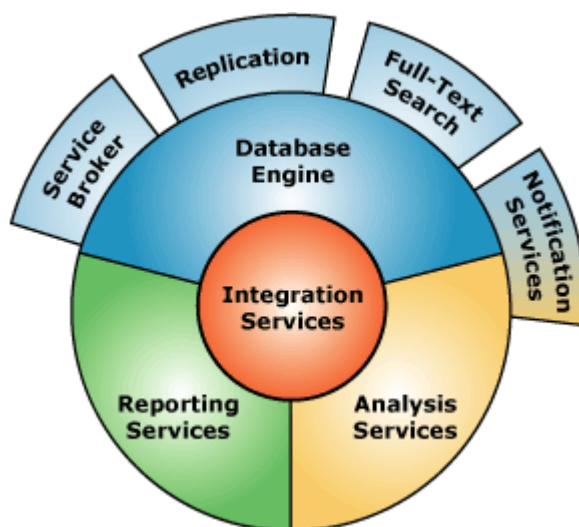
Obrázok č.1.: Logo spoločnosti, Zdroj: Intelligent technologies

Česká konzultační a vývojářská společnost Intelligent Technologies pomáhá svým klientům dělat lepší rozhodnutí. Společnost implementuje a dodává řešení v oblasti business intelligence a data warehousingu. [1]

1 Teoretické východiská práce

1.1 MS SQL Server

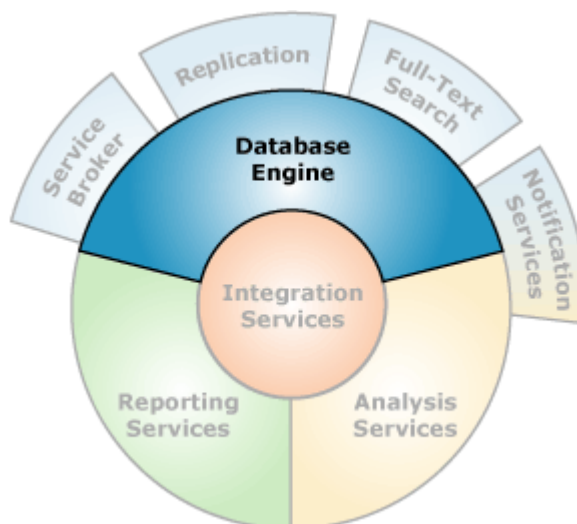
“Microsoft SQL server je databázová platforma pre veľké online transakčné procesenie (OLTP), dátové sklady a e-commerce aplikácie. Je to taktiež platforma BI pre integráciu, analýzu a reportovanie dát.” [7]



Obrázok č.2.: MS SQL Server, Zdroj: technet.microsoft.com

1.2 MS SQL Database engine

“Microsoft SQL Server databázový engine je základná služba pre ukladanie, procesenie a zabezpečovanie dáta. Databázový engine poskytuje kontrolovaný prístup a rýchle transakčné procesenie aby pokrylo minimálne požiadavky aplikácií, ktoré potrebujú veľké množstvá dát v rámci spoločnosti. Databázový engine taktiež poskytuje bohatú podporu pre udržiavanie vysokej prístupnosti.” [7]



Obrázok č.3.: MS SQL Database Engine, Zdroj: technet.microsoft.com

1.3 MS SQL SRS

“Od roku 2003, keď mali užívatelia poprvý krát možnosť pracovať s reportovacími službami, sa spôsob akým sa tento produkt používal sa produkt samotný podstatne zmenil. Namiesto jednoduchého nástroja na vytváranie tabuľkových reportov, sa stal základom na kotrom je možné konštruovať kompletne reporty, skórovacie karty a správy o pokrokoch pre podnikových používateľov a konzultácie so zákazníkmi. Toto nie je reportovací program vášho otca. Dnes dokáže všetko od jednoduchých ad hoc dátových reportov až po doručovanie enterprise-ready, integrovaného reportingu na biznis portály alebo vlastné aplikácie. Teraz, v tretom vydaní pre SQL Server 2008, sa reportovacie prostredie dramaticky zlepšilo, s veľmi dobre upravenou produktovou architektúrou a prídavkom niekoľkých zlepšených funkcií. Reportovacie služby neprichádzajú len s preukázateľnými výsledkami ale sú taktiež schopné pokryť potreby vážnych IT developerov a biznis informačných pracovníkov, ktorí potrebujú len jednoduchý nástroj bez komplexnej a technickej sofistikovanosti.” [2]

Obrázok č.4.: „*This is not your father’s reporting tool.*“



Obrázok č.5.: MS SQL Server Reporting Services, Zdroj:

technet.microsoft.com

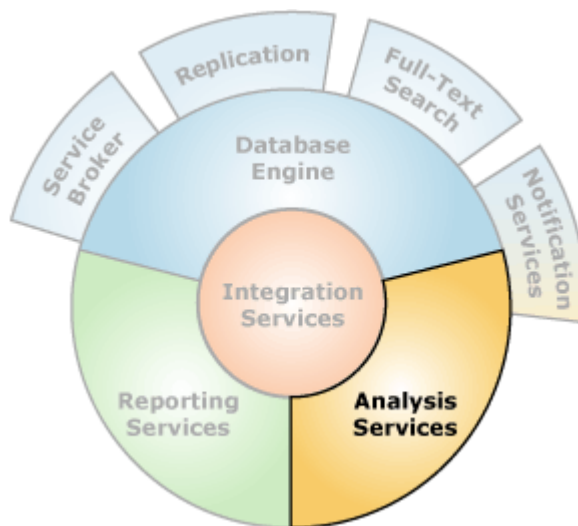
Kto používa reportovacie služby?

Pravdepodobne jedna z najvýznamnejších lekcí za posledných 5 rokov učenia tréningových kurzov s reportovacími službami je to aká je rôznorodosť publika. Role a pozadie tých, ktorí navrhujú a implemetnujú reportovacie riešenia sú rozdielne. Ako vývojár aplikácií som bol zvyknutý učiť programátorov a iných technologických profesionálov, ktorých životný účel je vytvoriť na zemi lepšie miesto a to písaním programového kódu a softvéru. [2]

V skratke povedané, MS SQL Server Reporting Services je nástroj na vytváranie vlastných reportov ktoré môžu mať podobu jednoduchej tabuľky alebo integrovaného firemného nástroja vo forme aplikácie alebo webovej stránky.

1.4 MS SQL SSAS

„Microsoft SQL Server Analysis Services (SSAS) poskytuje online analytické procesy (OLAP) a funkcionality dátového miningu pre aplikácie business intelligence. Analytické služby podporujú OLAP tým, že je v nich možné navrhnuť, vytvoriť a spravovať multidimenzionálne štruktúry, ktoré obsahujú agregované dáta z iných dátových zdrojov, ako sú napríklad relačné databázy. Pre aplikácie dátového modelovania umožňujú analytické služby navrhnuť, vytvoriť a vizualizovať modely dátového modelovania, ktoré sú skonštruované z iných dátových zdrojov pomocou veľkej škály algoritmov dátového modelovania v priemyselnej oblasti.“ [6]



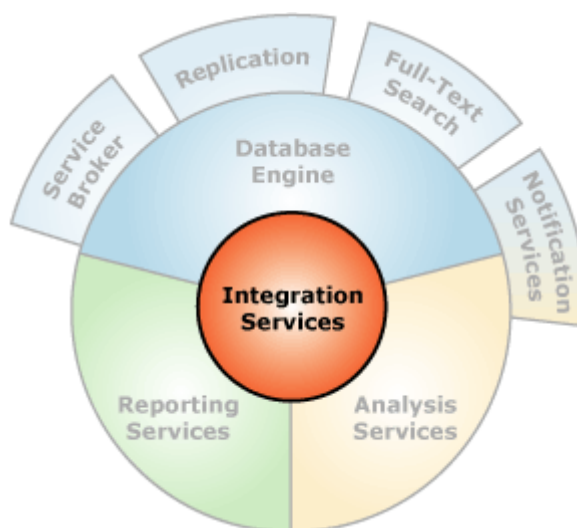
Obrázok č.6.: MS SQL Server Analysis Services, Zdroj:
technet.microsoft.com

1.5 MS SQL SSIS

„Microsoft Integration Services (integračné služby spoločnosti Microsoft) je platforma pre budovanie veľkopodnikových dátových integračných a transformačných riešení. Integračné služby sa využívajú na riešenie komplexných biznis problémov, ktoré môžu vzniknúť kopírovaním alebo sťahovaním dát, posielaním e-mailov ako odpovedí

na pozvánky, updatovaním dátových skladov, čistením a minovaním dát, a spravovaním SQL Serverových objektov a dát. Balíčky dokážu pracovať samostatne alebo spolu s inými balíčkami aby pokryli zložité biznisové potreby. Integračné služby vedia vytiahnuť a transformovať dáta z veľkej škály vstupov ako napríklad XML dátové súbory, flat file súbory (flat file súbory obsahujú dáta, ktoré nemajú štrukturované vzťahy) a relačné dátové zdroje a potom ich načítať do jednej alebo viacerých destinácií.” [5]

„Integračné služby obsahujú veľkú škálu vstavaných úloh a transformácií (nástrojov na vytváranie balíčkov) a službu integračných služieb pre beh a menežovanie balíčkov. Je možné použiť grafické prostredie integračných služieb bez toho aby bolo nutné napísať čo i len jediný riadok kódu alebo je možné naprogramovať rozsiahly objektový model integračných služieb pre vytváranie balíčkov programovo a nakódovať tak vlastné úlohy a iné objekty balíčkov.” [5]



Obrázok č.7.: MS SQL Server Integration Services, Zdroj:
technet.microsoft.com

„SQL Server Integretion Services bolo vypustené na trh ako súčasť SQL servera 2005 a tak prekvapila trh s ETL. V SQL server 2008 sa integračné služby zamerali na dospievanie produktov a zlepšovanie škálovetelnosti produktov v niektorých prípadoch

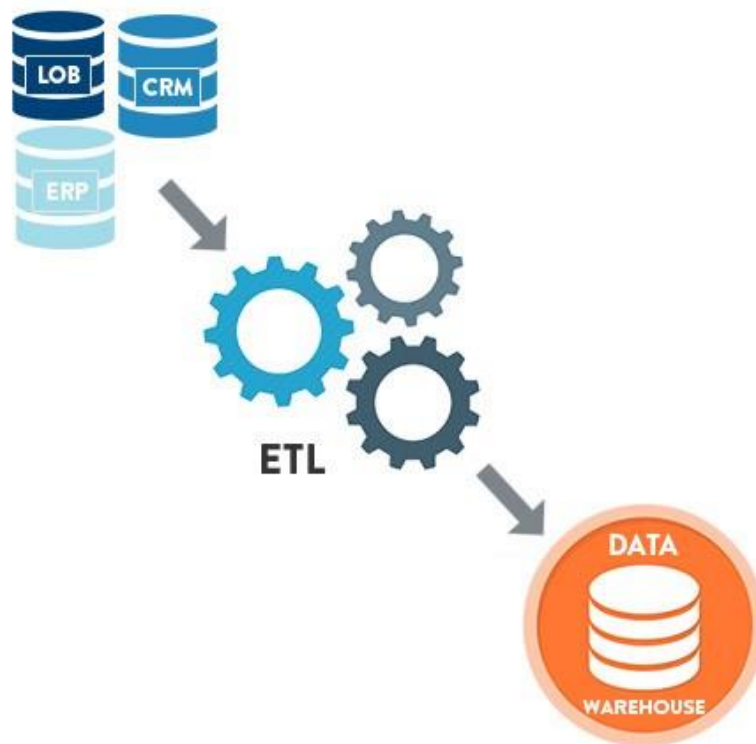
o ohromujúcich 70%. Ak ste nový v SSIS, tak ste si vybrali výbornú oblasť, v ktorej začať. Jediná potrebná zručnosť, ktorú je treba vedieť na dnešnom technickom trhu práce je ETL.“ [3]

1.6 Business Intelligence

„Business intelligence (BI) zjednodušuje objavovanie informácií a ich analýzu, pomáha robiť rozhodnutia pracovníkom na kažej úrovni v organizácii a to ľahším prístupom, porozumením, analýzou, spoluprácou a prácou s informáciami kedykoľvek a kedykoľvek. Táto definícia pre BI demonštruje to, že sa tradičné analýzou riadené BI aplikácie vyvinuli tak aby obsahovali niekoľko iniciatív na meranie, riadenie a zlepšenie výkonnosti jednotlivcov, procesov, tímov a obchodných jednotiek. Kvôli tomu, Microsoft ponúka BI nástroje, ktoré umožňujú každému zamestnancovi prístup k dátam potrebným k tomu aby robili informované rozhodnutia, a tak mali možnosť pracovať tak ako to vedia.” [8]

1.7 ETL

*„Extrakcia, transformácia a načítanie (**E**xtraction, **T**ransformation, **L**oading) systém DW/BI prostredia pozostáva z pracovnej sféry, dátových štruktúr a skupín procesov. ETL systém je všetko medzi operačnými zdrojovými systémami a DW-/BI prezentačnou oblasťou.“*



Obrázok č.8.: ETL proces, Zdroj: images.google.com

„Extrakcia je prvým krokom v procese vkladania dát do prostredia dátového skladu. Extrakcia znamená čítanie a porozumenie zdrojových dát a kopírovanie potrebných dát do ETL systému pre ďalšiu manipuláciu. V tomto bode dáta patria do dátového skladu.”

„Po tom ako sú dáta extrahované do ETL systému, prejdú množstvom potenciálnych transformácií, ako napríklad čistenie dát (oprava chýb, riešenie doménových konfliktov, popasovanie sa s chýbajúcimi prvkami alebo zjednotenie formátu dát), kombinovanie dát z viacerých zdrojov a očistenie dát od duplicít. ETL systém pridáva hodnotu dátam pomocou týchto čistiacich a prispôbovacích úkonov tým, že mení dáta a vylepšuje ich. Navyše tieto aktivity môžu byť postavené na to aby vytvárali metadáta pre diagnostiku poprípade môžu časom viesť k spätnému vývoju, alebo úprave biznisových procesov pre úpravu kvality dát v zdrojových systémoch.”

„Posledným krokom v ETL procese je fyzické štruktúrovanie a načítanie dát do prezentačnej oblasti cieľových dimenzionálnych modelov. Pretože hlavnou misiou ETL systémov je predať tabuľky dimenzií a tabuľky faktov v dodacom kroku, sú tieto podsystémy kritické. Veľa týchto podsystémov sa zameriava na spracovanie tabuliek dimenzií, ako napríklad priradenie náhradných (umelých) kľúčov, dohľadávanie kódov pre pripojenie vhodného opisu, rozdeľovanie alebo kombinovanie stĺpcov pre zobrazenie vhodných dátových hodnôt alebo spájanie základných tabuľkových štruktúr, ktoré sú v tretej normálovej forme, do plochých denormalizovaných dimenzií.“ [9]

1.8 Data flow task

„Úloha toku dát zapuzdruje “motor“ na prenos údajov, ktorý presúva dáta medzi zdrojmi a cieľmi, a umožňuje používateľovi transformovať, čistiť a upravovať údaje pri ich presune. Pridanie úlohy dátového toku do riadiaceho toku obalu umožňuje balíčku získavať, transformovať a načítať dáta.“

„Tok údajov pozostáva aspoň z jednej zložky toku údajov, ale zvyčajne je to súbor pripojených dátových tokov: zdroje, ktoré extrahujú dáta; transformácie, ktoré menia, smerujú alebo sumarizujú údaje a ciele, ktoré ukladajú údaje.“

„Úloha toku údajov môže zahŕňať viac tokov údajov. Ak úloha kopíruje viacero sád dát a poradie, v ktorom sa údaje kopírujú, nie je významné, môže byť vhodnejšie zahrnúť viac tokov údajov do úlohy toku dát. Môžete napríklad vytvoriť päť dátových tokov, každé kopírované dáta z “plochého“ súboru do inej tabuľky dimenzií vo hviezdicovej schéme dátového skladu.“ [7]

1.9 Tabuľky dimenzií

„Tabuľky dimenzií sú integrovanými spoločníkmi k tabuľke faktov. Tabuľky dimenzií obsahujú slovný popis spojený s biznisovými procesmi, ktoré ohodnocujú výkonové ukazatele. Nie je nezvyklé pre tabuľku dimenzií aby mala 50 až 100 atribútov,

ale nájdú sa aj také, ktoré ich majú len pár. Tabuľky dimenzií majú poväčšine menej riadkov ako tabuľky faktov avšak sú omnoho širšie s veľkým množstvom textových stĺpcov. Každá dimenzia je definovaná jediným primárnym kľúčom, ktorý odkazuje do tabuľky faktov.” [9]

1.10 Tabuľky faktov

“Tabuľka faktov v dimenzionálnom modeli uchováva výkonové ukazatele plynúce z biznisových procesov spoločnosti. Pretože dáta, ktorými sa ohodnocuje, sú ohromne veľké množstvá dát, nemali by byť replikované na niekoľko miest pre viac funkcií organizácie naprieč celou spoločnosťou.” [9]

1.11 OLAP

“OLAP (Online Analytical Processing) je technológia, pomocou ktorej možno organizovať veľké podnikové databázy a podporovať biznisovú inteligenciu. Databázy OLAP sú rozdelené do jednej alebo niekoľkých kociek, pričom každá z nich je organizovaná a navrhnutá správcom kocky tak, aby vyhovovala vlastnému spôsobu získavania a analyzovania údajov, a aby uľahčila vytváranie potrebných zostáv kontingenčnej tabuľky a kontingenčného grafu.”

“Databázy OLAP (Online Analytical Processing) uľahčujú riešenie dotazov týkajúcich sa biznisovej inteligencie. OLAP je databázová technológia optimalizovaná skôr na vytváranie dotazov a zostáv údajov než na spracovávanie transakcií. Zdrojovými údajmi pre technológiu OLAP sú databázy OLTP (Online Transactional Processing), ktoré sú zvyčajne uložené v dátových skladoch. Údaje OLAP sú odvodené z historických údajov a agregované v štruktúre, ktorá umožňuje ich vysoko špecializovanú analýzu. Tieto údaje sú organizované hierarchicky a nie sú uložené v tabuľkách, ale v kockách. Táto sofistikovaná štruktúra poskytuje rýchly prístup k údajom určeným na analýzu pomocou viacrozmerných štruktúr. Vďaka tejto organizácii údajov možno jednoducho zobrazit špecializované zhrnutia údajov pomocou zostavy

kontingenčnej tabuľky alebo kontingenčného grafu, napríklad tržby v rámci celej krajiny alebo oblasti alebo zobraziť podrobnosti o lokalitách s mimoriadne vysokými alebo nízkymi tržbami.” [7]

1.12 Jazyk SQL

„SQL je databázový jazyk navrhnutý pre získavanie a riadenie dát v relačných databázových riadiacich systémoch (DBMS), pre vytváranie a modifikácie základných prvkov databázy, a pre riadenie prístupu k databázovým objektom. Skratka SQL znamená Structured Query Language.

Je odvodený zo svojho predchodcu - jazyka SEQUEL, ktorý bol vyvinutý IBM začiatkom sedemdesiatych rokov. Pôvodne bol navrhnutý ako tzv. deklaratívny jazyk, v súčasnosti je rozširovaný o prvky tzv. príkazových (imperatívnych) jazykov, ako sú napríklad príkazy toku riadenia (if-then-else).

Pod pojmom "deklaratívny jazyk" sa zvyčajne myslí, že jazyk len "deklaruje" ako má vyzerat' výsledok a nedefinuje algoritmus, ktorým sa výsledok má dosiahnuť. V prípade dotazov jazyka SQL ide o špecifikáciu podmienok, ktorú má spĺňať výstupná množina.

Alternatívne, pojem "deklaratívny jazyk" označuje, že ide o jazyky, ktorý používajú postupy funkcionálneho, logického resp. "constraint" programovania.“ [6]

1.13 Jazyk MDX

“Ako je SQL jazyk pomocou ktorého vyťahujeme dáta z relačných databáz, tak MDX (Multi-Dimenzionálne výrazy (eXpressions)) je jazyk používaný na výťah dát z multidimenzionálnych databáz. Presnejšie, MDX sa používa na dopytovanie mulitidimenzionálnych dát z analytických služieb a podporuje dva rozdielne módy. Keď

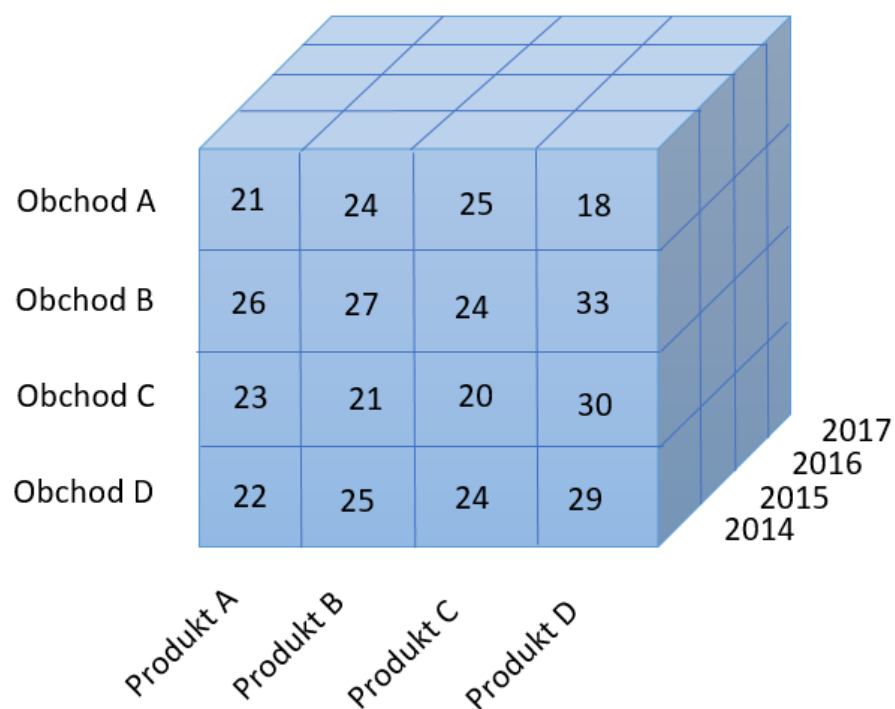
sa použije vo výraze, MDX vie definovať a manipulovať multidimenzionálne objekty a dáta tak aby vypočítal hodnoty. Ako dopytovací jazyk sa používa na výber dát z databázií analytických služieb. MDX jazyk bol pôvodne navrhnutý spoločnosťou Microsoft a predstavený v SQL Server Analysis Services 7.0 v roku 1998.

Jazyk MDX nie je vlastnený. Je to štandardizovaný dopytovací jazyk používaný pre výber dát z OLAP databázií. MDX je časťou OLE DB pre OLAP špecifikácie sponzorované spoločnosťou Microsoft. Mnoho iných OLAP poskytovateľov podporuje MDX, vrátane Intelligence server spoločnosti Microstrategy, Essbase server od Hyperionu a Enterprise BI server od spoločnosti SAS. Existuje taktiež mnoho ľudí, ktorí by chceli aby mal jazyk MDX ďalšiu funkcionálnu a tak sú rozšírenia pre MDX jazyk vyvíjané individuálnymi obchodníkmi.” [4]

1.14 Multidimenzionálna databáza

“Multidimenzionálna databáza je zvyčajne označovaná ako kocka (Cube). Kocka je základ multidimenzionálnej databázy a každá kocka bežne obsahuje viac ako dve dimenzie. Napríklad kocka v ukážkovej databáze Adventure Works spoločnosti Microsoft obsahuje 21 dimenzií.” [4]

Najbežnejší spôsob ako si predstaviť (popísať vysvetliť) multidimenzionálnu databázu je tak že si ju je možné predstaviť ako už spomínanú kocku, na ktorej hranách sú jednotlivé dimenzie a obsah kocky tvoria fakty. Napríklad kocka predajov. Na jednej hrane sú dimenzie produktov, na druhej hrane dimenzie obchodov a na tretej sú dimenzie s dátumovými atribútmi. Takáto kocka nám ukazuje v aké produkty boli predané v určitom obchode a v daný čas a sumu za ktorú boli predané.



Obrázok č.9.: Dátová kocka, Zdroj: Vlastná tvorba

1.15 Data warehouse

“Dátový sklad porporuje až 2 terabajty dát, ktoré majú vlastné časové známky kvôli sledovaniu zmien. Skladisko dát je vytvorené pomocou automatických synchronizácií z konfiguračného menežera lokálných databáz do databázy dátového skladu. Tieto informácie sú potom prístupné z reportovacích služieb.” [7]

Analýza súčasného stavu

1.16 SLEPT analýza – analýza obecného okolia

1.16.1 Sociálne faktory

Medzi hlavný sociálny faktor ovplyvňujúci firmu je predovšetkým viditeľnosť pre zákazníkov, ktorí požadujú outsourcovanie časti svojich IT riešení v oblasti reportingu a data warehousingu. Momentálne má firma niekoľko svojich zákazníkov, ktorým ponúka podporu pre svoje riešenia a taktiež má v ponuke množstvo firiem s ktorými môže podpísať zmluvy o spolupráci pokiaľ si ju vyberú spomedzi množstva konkurencie. Výhodou pre firmu je to, že sa zviditeľňuje na viacerých konferenciách, kde aj prednáša svoje už hotové riešenia a takisto aj referencie od zákazníkov, ktorí sú spokojní s vytvorenými riešeniami.

1.16.2 Legislatívne faktory

Keďže sa jedná o plne legálnu firmu tak musí splňovať niekoľko zákonov a právnych noriem Českej republiky. Napríklad vyhláška o podvojnom účtovníctve pre firmy, zákon o zodpovednosti zamestnávateľa a pod. Taktiež firma uzatvára zmluvu o mlčanlivosti, ktorá chráni zákazníka pred únikom citlivých dát, keďže firma pracuje práve s týmito aktívami.

1.16.3 Ekonomické faktory

Problémy ekonomického charakteru by nemali byť momentálne akokoľvek ohrozujúce existenciu firmy, avšak ak sa Česká národná banka rozhodne hýbať s kurzom českej koruny tak to môže ovplyvniť vzťahy so zákazníkmi mimo Českej republiky. Ďalším rizikom budúcnosti či už kladným, alebo záporným môže byť finančná resp. ekonomická situácia ostatných firiem, ktoré pri väčšej záťaži môžu

znižovať potreby firmy tým, že sa zamerajú iba na hlavné činnosti. To znamená, že nebudú požadovať reporting svojich aktivít a tým pride firma o zákazníkov. Na druhú stranu, kladným faktorom môže byť opak, a to že sa firmám darí po ekonomickej stránke a majú financie na ostatné podporné aktivity.

1.16.4 Politické faktory

Problémy politického faktoru by taktiež nemali byť v tejto dobe veľmi výrazné, avšak tu tiež existuje riziko zmeny, napr. sa môže zmeniť daňová záťaž na podnikateľské subjekty.

1.16.5 Technologické faktory

Pri odetví, v ktorom firma funguje je potrebné aby sledovala nové trendy v oblasti IT technológií. Problémom tu môže nastať to, že zákazníci nemajú vždy najnovšie produkty ktoré by uľahčili vývoj nových riešení a to pre to, že by si museli kúpiť nové verzie stávajúcich softwarových produktov.

1.17 Porterov model piatich konkurenčných síl – analýza oborového okolia

1.17.1 Konkurencia prostredia

Na trhu v danom odvetví nie je veľké množstvo konkurenčných firiem, ktoré by spoločnosť vo väčšej miere ohrozovali. Navyše je firma Intelligent Technologies partnerom spoločnosti Microsoft, a tak má oproti ostatným v niečom navrch. Samozrejme nie je jediná, ktorá je partnerom tejto spoločnosti avšak podobých firiem je na českom trhu málo.

1.17.2 Hrozba substitútov

V tomto odvetví neexistuje veľká hrozba substitútov, keďže riešenia, ktoré firma dodáva su ušité na mieru presne podľa požiadaviek zákazníka.

1.17.3 Zákazníci

Zákazníci spoločnosti Intelligent Technologies sú častejšie väčšie firmy, ktoré sa snažia mať prehľad vo svojich aktivitách ale nechcú sa tým sami zaťažovať. Preto tieto firmy používajú riešenia, ktoré im firma poskytne a to za nižšie časové a peňažné náklady.

1.17.4 Dodávatelia

Hlavným dodávateľom je spoločnosť Microsoft, ktorá sama vyvíja prostredia a programy umožňujúce vývoj produktov, ktoré od firmy zákazníci požadujú.

1.17.5 Riziko vstupu nových konkurentov

Riziko vstupu nových konkurentov je v dobe informačných technológií podstatne vyššie ako to bývalo v minulosti, avšak firma už má svoje vybudované meno a riziko nových konkurentov nie je v tomto prípade príliš vysoké na to aby sa firma bála o svoju existenciu a svojich ako stávajúcich tak aj nových zákazníkov.

1.18 Model 7S – analýza interných faktorov

1.18.1 Stratégia

Dlhodobou stratégiou firmy je to aby držala krok so súčasným technologickým vývojom aby mohla svojim zákazníkom ponúknuť tie najlepšie možné riešenia. To ale znamená sledovať nové trendy a učiť sa nové veci, ktoré tieto trendy so sebou prinášajú.

1.18.2 Štruktúra

Štruktúra vo firme je lineárna na čele s dvomi majiteľmi.

1.18.3 Informačné systémy

Firma má svoj zaužívaný systém, v ktorom sa jednotlivé veci evidujú. Ide o záznamy v programe excel, onenote, trello ai. Hlavným plusom je, že firma získala certifikát ISO 9001 a to znamená zavedenie systému kontroly kvality (QMS), ktorý definuje procesy podľa ktorých sa spoločnosť riadi pri prácach na projektoch.

1.18.4 Štýl riadenia

Riadenie vo firme je projektové a neexistuje manažér manažéra. Každý môže do diskusie priniest vlastné nápady.

1.18.5 Spolupracovníci

Zamestnancov firmy tvoria pracovníci so skúsenosťami v oblasti BI systémov ale aj nováčikovia, ktorým je však poskytnutá možnosť ďalšieho vzdelávania v oblastiach, ktoré sa daného oboru týkajú. Vo firme panuje uvoľnená a rodinná atmosféra ako medzi jednotlivými pracovníkmi tak aj voči vedeniu.

1.18.6 Zdieľané hodnoty

Zdieľané hodnoty súvisia so spolupracovníkmi a samotnou firemnou kultúrou. Na druhú stranu sa firma snaží robiť všetko pre to aby bola korektná ako voči svojim zákazníkom, tak aj k svojim zamestnancom a okoliu.

1.18.7 Schopnosti

Schopnosti firmy sa pohybujú na vysokej úrovni. Ako už bolo spomenuté, firma je partnerom spoločnosti Microsoft a získala od nej množstvo ocenení. Tak isto sa previdelne zúčastňuje na rôznych IT konferenciách, kde niekedy aj prednáša.

1.19 SWOT analýza

1.19.1 Silné stránky

- **Certifikovaný partner spoločnosti Microsoft**

Firma Intelligent Technologies s.r.o. je certifikovaným partnerom spoločnosti Microsoft ako jedna z mála v Českej republike. Znamená to, že sa zúčastňuje na jej konferenciách a naplno využíva jej produkty pri svojej práci.

- **Certifikácia ISO 9001**

Od 6 Februára 2017 firma získala certifikát ISO 9001:2008, čo znamená, že je vo firme zavedený system kontroly QMS (quality management system). Čiže všetko čo sa vo firme robí musí byť definované procesmi, ktoré to všetko riadia. Malo by sa tým zamedziť situáciám, že každý projekt je iný a tým vynikne chaos a nerodrozumenie pri práci naprieč projektami, a že každý zákazník dostane inú službu. V prípade Intelligent Technologies to znamená, že nový zákazník je založený v CRM, je mu ponúknutá zmluva o ochrane informácií a rámcová zmluva o spolupráci, vznikne interná skupina, požiadavky su registrované a práca je detailne evidovaná. To znamená, že všetky projekty by mali byť vypracované podľa rovnakého vzoru a taktiež aj do administratívy.

- **Uvoľnená atmosféra pre zamestnancov**

Zamestnanci sa v práci cítia pohodlne vďaka tomu, že vo firme panuje príjemná a rodinná atmosféra. To vďaka tomu, že je firma relatívne malá.

- **Silné meno v obore**

Svojou prácou si firma vyslúžila silnú pozíciu na trhu a to ako účasťou na medzinárodných konferenciách tak aj spätnou väzbou zo strany svojich zákazníkov.

- **Prisposobenie sa potrebám zákazníkov**

Každý zo zákazníkov je jedinečný a tým sú aj ich požiadavky. Pre to sa pri práci na riešeníach po celú dobu komunikuje so zákazníkom čím sa zamedzia prípadné nezhody v tom ako si produkt zákazník predstavoval a v tom čo mu nakoniec bolo doručené.

- **Účasť na konferenciách**

Už vyššie spomínaná súťaž na medzinárodných konferenciách je prostriedkom ako sa zviditeľniť a získať tak nových zákazníkov a to aj mimo územia Českej republiky.

1.19.2 Slabé stránky

- **Firma je dosť malá a v prípade väčšieho záujmu nemusí byť schopná pokryť dopyt zákazníkov**

Počtom zamestnancov firma Intelligent Technologies predsaden nesiahá na vysoké priečky v rebríčkoch počtov zamestnancov. Pre to sa môže stať, že nebude schopná pokryť väčší dopyt po jej službách a produktoch.

- **Špecializácia iba na Microsoft**

Aj keď je sú produkty a platformy spoločnosti Microsoft vo veľkom množstve používané, nie vždy si ju firmy vyberajú ako svoje pracovné prostredie. To znamená, že pri zameraní na platformy spoločnosti Microsoft nie je firma schopná získať zákazníkov z iných oblastí resp. platforiem.

- **Nedostatočná spätná väzba zo strany zákazníkov**

Aj keď zákazníci spolupracujú pri tvorbe riešení, nie vždy sú schopní presne povedať ako sa dané riešenie bude u nich používať. Pre to firma nevie či sa v jej produktoch nevyskytujú malé problémy, ktoré môžu byť časovo alebo výpočetne nákladné a tak nevie presne upraviť riešenie aby malo čo najmenšiu záťaž či už na čas alebo na stroje.

1.19.3 Príležitosti

- **Stále viac firiem potrebuje mať prehľad o svojích aktivitách**

V modernej dobe je trendom outsourcovať interné potreby a preto si firmy nechávajú vypracovať prehľady iným firmám ako je napríklad Intelligent Technologies. Robia tak pre to, že sa snažia zamerať na kľúčové prvky a na ostatné veci nechcú míňať svoje zdroje, hlavne financie. Pre to je pre zákazníka finančne a časovo menej náročné nechať si vypracovať prehľad o svojom podnikaní druhou firmou, ktorá sa tým špecializuje.

- **Industry 4.0 – očakávaný vysoký nárast dát**

Ďalším trendom je kompletná digitalizácia procesov vo firme. Tým pádom sa zvyšuje objem dát, ktoré sa ukladajú a ktoré je dobré následne analyzovať pre zistenie výskytu možných nedostatkov alebo problémov, ktoré môžu byť pre firmy niekedy veľmi nákladné.

1.19.4 Hrozby

- **Strata zákazníkov voči konkurencii**

Aj keď ma firma Intelligent Technologies svojich loajálnych zákazníkov a silne vybudované meno na trhu, vždy je tu šanca, že sa zákazníci začnú obzerať po niečom inom. S tým súvisí aj ďalšia hrozba vstupu nových konkurentov na trh.

- **Vstup nových konkurentov na trh**

Keďže sa očakáva vysoký nárast dát s príchodom Industry 4.0 je dosť možné, že sa na trhu s Business Intelligence a Data Warehousing objavia nové spoločnosti ponúkajúce svoje produkty. To znamená hrozbu nových konkurentov a stratu zákazníkov voči nim.

- **Strata záujmu zo strany zákazníkov**

Ekonomický cyklus môže dostať firmy do takeého stavu, že sa budú musieť vzdať okrajových procesov aby si mohli udržať procesy kľúčové. To pre firmu znamená stratu zákazníkov a tým pádom zníženie prílevu finance do firmy.

- **Microsoftu sa môže prestať dariť**

V neposlednom rade tu existuje možnosť, že spoločnosť Microsoft prestane byť pre zákazníkov príťažlivá oni prejdú k produktom konkurencie, poprípade k opensource riešeniam. To by taktiež znamenalo pokles dopytu pro riešeniach spoločnosti Intelligent technologies.

1.20 Záver analýzy

Z analytickej časti je zrejmé, že firma má svojich loajálnych zákazníkov a má aj dostatočný podiel na trhu v oblasti BI riešení. Nie je vylúčené, že sa na trh nedostanú nové spoločnosti ponúkajúce podobné produkty, ale aj v tom prípade má spoločnosť dobre postavené meno v obore.

Zo slabých stránok firmy je vidieť, že sa nejdená o veľký a ani stredný podnik preto sa môže stať, že nemusí pokryť požiadavky svojich zákazníkov v prípade, že bude zvýšený dopyt po ich produktoch, a že nie vždy dostane od svojich zákazníkov dostatočne jasnú spätnú väzbu, podľa ktorej by svoje produkty mohla vylepšiť. Preto sa v nasledujúca časť práce zameriava práve na to.

2 Vlastný návrh riešenia

2.1 Popis navrhovaného riešenia

Hlavným účelom tejto práce je navrhnuť a vytvoriť riešenie, ktoré sa bude snažiť pokryť nedostatky zistené z analýzy spoločnosti. Malo by sa jednať o profilovacie a monitorovacie riešenie nad business intelligence systémami a malo by byť použiteľné na akýkoľvek projekt, resp. dátový sklad, ktorý firma Intelligent Technologies vytvorí.

Riešenie umožní sledovať spúšťanie reportov (možstvo, počet ľudí, dĺžka behu, jednotlivé reporty a zložky, ...), taktiež monitoring práce s kockou (kto a kde kocku používa, ktoré oblasti sú využité, ako dlho dotazy trvajú). Výsledkom riešenia bude OLAP kocka so všetkými týmito informáciami s možnosťou vytvorenia akýchkoľvek reportov nad touto kockou.

Hlavným prínosom riešenia by malo byť to, že sa bude dať nasadiť obecne na akýkoľvek SQL dátový sklad. Takto vytvorené riešenie a reporting nad ním uľahčí prácu vývojárom, ktorí budú pracovať na nových zákazkách tým že budú schopní sa zacieliť na to ako sa jednotlivé riešenia môžu používať na strane zákazníka a tak ich ešte lepšie prispôsobiť požiadavkom jednotlivých zákazníkov.

2.2 Lewinov model

2.2.1 Dôvod zmeny

Dôvodom zmeny je zhrnutie všetkých vyššie spomenutých vlastností do jedného celku aby sa vytvoril celkový pohľad na prácu s dátovými skladmi.

2.2.2 Agent zmeny

Agent zmeny je zamestnanec firmy, ktorý to má na starosti, samozrejme s podporou vedenia a skúsenejších spolupracovníkov.

2.2.3 Intervenčné oblasti

Intervenčnou oblasťou zmeny je hlavne prehľad o práci s vytvorenými riešeniami a podpora novovyvíjaných riešení.

2.2.4 Realizácia zmeny

Zmena bude realizovaná pomocou vlastných znalostí, znalostí spolupracovníkov a pomocou ďalších znalostí nadobnutými počas realizácie projektu.

2.2.5 Zhodnotenie zmeny

Novovytvorené riešenie by sa malo prejavovať na zlepšení kvality práce na jednotlivých produktoch ako pre stávajúcich tak aj pre nových zákazníkov.

2.3 Zber dát

Zber telemetrických dát, uložených v konfiguračnej databáze a ostatných dát o obsahu dátového skladu sa uskutočňuje pomocou niekoľkých ETL balíčkov v prostredí programu Visual Studio 2015. Dva z balíčkov boli vo firme už nejaký ten čas predpripravené avšak neboli dokončené a ani nikdy použité pre účely aké sa nimi táto práca snaží docieľiť. Jedná sa o balíčky “Load Report Log” a “Load Query Log”, ktoré sú podrobnejšie popísané ďalej. Zvyšné ETL balíčky sú vytvorené podľa potreby a doteraz sa nikde nevyskytovali. Zbierané dáta sa ukladajú do relačnej databázy, ktorú je

nutné si vytvoriť ešte pred tým než sa začne pracovať s jednotlivými ETL procesmi pretože by dochádzalo k chybám spojenia.

2.3.1 OLAP Query Log

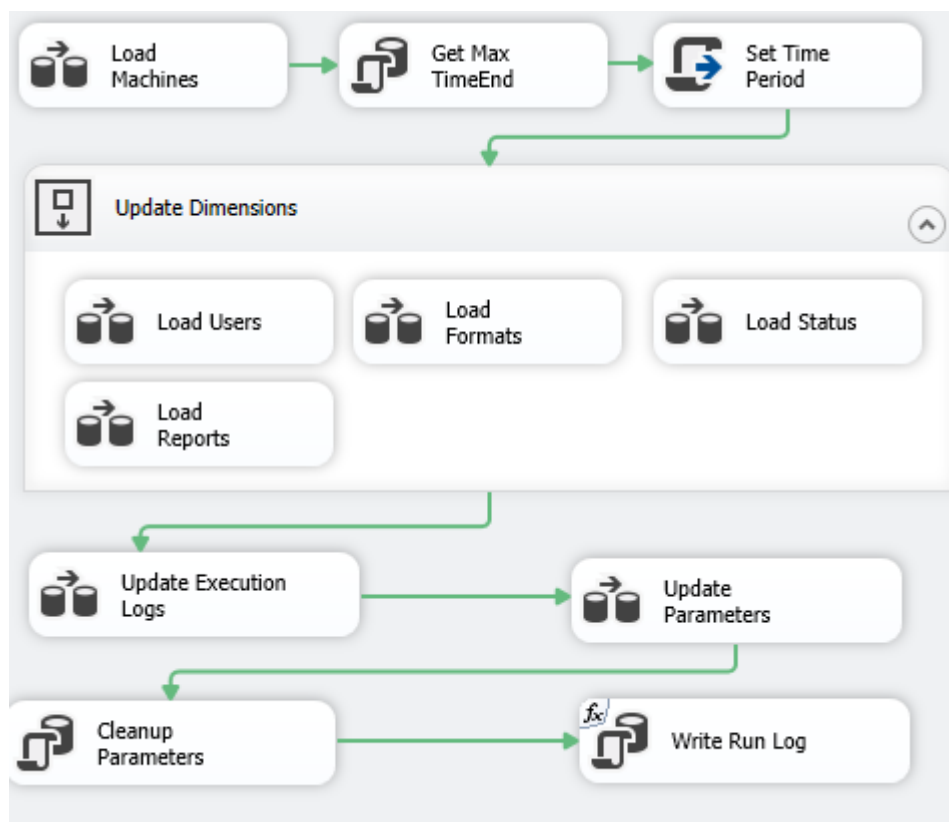
Na začiatok zberu dát bolo potrebné vytvoriť, resp. zapnúť automatické logovanie dát do tabuľky. Preto bolo nad multidimenziálnou databázou nastavené zapisovanie do tabuľky “OlapQueryLog” kam sa zapisovali údaje o užívateľovi, databáza nad ktorou bol daný dataset spustený, samotný dataset v binárnom zápise, dĺžka trvania procesu, datum a čas začatia procesu a cesta k objektu či už ide o report alebo len SQL dotaz.

| | MSOLAP_Database | MSOLAP_ObjectPath | MSOLAP_User | Dataset | StartTime | Duration |
|---|-----------------|--------------------------------|-----------------|--|-------------------------|----------|
| 1 | Delmart | DWH.Delmart.Prodej.Prodej | DELMART\delmart | 0000000.000101000000100000.000000000000000000000000... | 2017-05-15 08:44:44.000 | 510 |
| 2 | Delmart | DWH.Delmart.Prodej.Prodej | DELMART\delmart | 0000000.000101000000100000.000000000000000000000000... | 2017-05-15 08:44:44.000 | 510 |
| 3 | Delmart | DWH.Delmart.Prodej.Prodej | DELMART\delmart | 0000000.000000000000000000.010000000000000000000000... | 2017-05-11 12:53:56.000 | 313 |
| 4 | Delmart | DWH.Delmart.Prodej.Prodej | DELMART\delmart | 0000000.000000000000000000.000000000000000000000000... | 2017-05-11 13:10:07.000 | 250 |
| 5 | Delmart | DWH.Delmart.Prodej.Prodej | DELMART\delmart | 0000000.000000000000000000.000000000000000000000000... | 2017-05-11 14:01:35.000 | 242 |
| 6 | Delmart | DWH.Delmart.Prodej.Prodej | DELMART\delmart | 0000000.000000000000000000.010000000000000000000000... | 2017-05-11 12:53:53.000 | 234 |
| 7 | Delmart | DWH.Delmart.Prodej.Prodej | DELMART\delmart | 0000000.000000000000000000.000000000000000000000000... | 2017-05-11 13:10:07.000 | 222 |
| 8 | Delmart | DWH.Delmart.Prodej.Fact Odpisy | DELMART\delmart | 0000000.000000000000000000.00000.0000.000000... | 2017-05-15 09:56:42.000 | 222 |
| 9 | Delmart | DWH.Delmart.Prodej.Prodej | DELMART\delmart | 0000000.000000000000000000.000000000000000000000000... | 2017-05-11 13:10:04.000 | 220 |

Obrázok č.10.: Olap Query Log Tabuľka, Zdroj: Vlastná tvorba

2.3.2 Load Report Log

Tento balíček slúži na zápis dát z používania jednotlivých reportov.

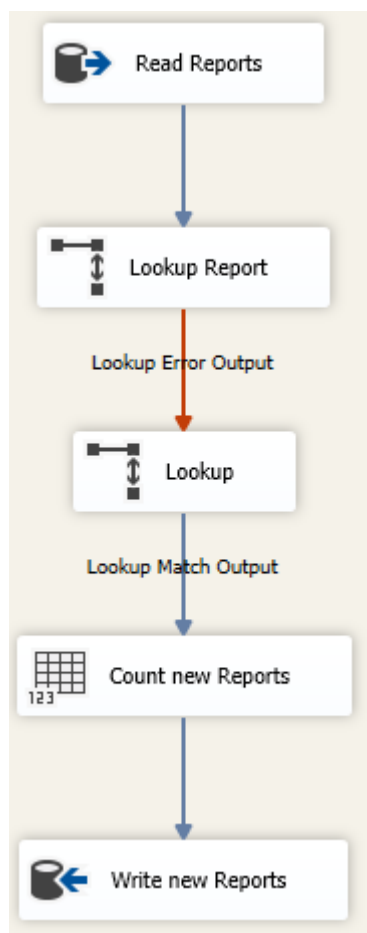


Obrázok č.11.: ETL pre využívanie reportov, Zdroj: Vlastná tvorba

Prvým krokom je dátový tok (Data Flow Task) pre načítanie serverov. Ten porovnáva stávajce a nové instance serverov pomocou “Lookup” úlohy a pokiaľ nájde nejaký nový server, ktorý nie je zapísaný v databáze tak daná úloha vyústi do chybného výstupu tak daný počet serverov doplní.

Ďalším data-flow taskom je zistenie maximálneho času ukončenia (MaxTimeEnd). Výstup z tejto úlohy postupuje ako vstup do ďalšej úlohy v ktorej sa nájdu dáta, ktoré majú časový príznak väčší ako zistený “MaxTimeEnd”. Ak sa v zdrojovej databáze take dáta nájdu tak sa zapíšu ako novo pridané.

Po skončení procesu zápisu nových dát sa paralelne začnú updatovať dáta v tabuľkách užívateľov, formátov, statusov a reportov. Všetky tieto úlohy majú rovnakú štruktúru ako má data-flow task pre načítanie serverov. To znamená načítanie dát zo zdrojovej databázy, porovnanie s dátami v cieľovej databáze, spočítanie nových výskytov a následne ich zápis.



*Obrázok č.12.: Úloha toku dát pre načítanie reportov z katalógu, Zdroj:
Vlastná tvorba*

Pri načítaní užívateľov je navyše jeden script napísaný v jazyku visual basic, ktorý nájde meno a department nových užívateľov pokiaľ nie sú umiestnené v active

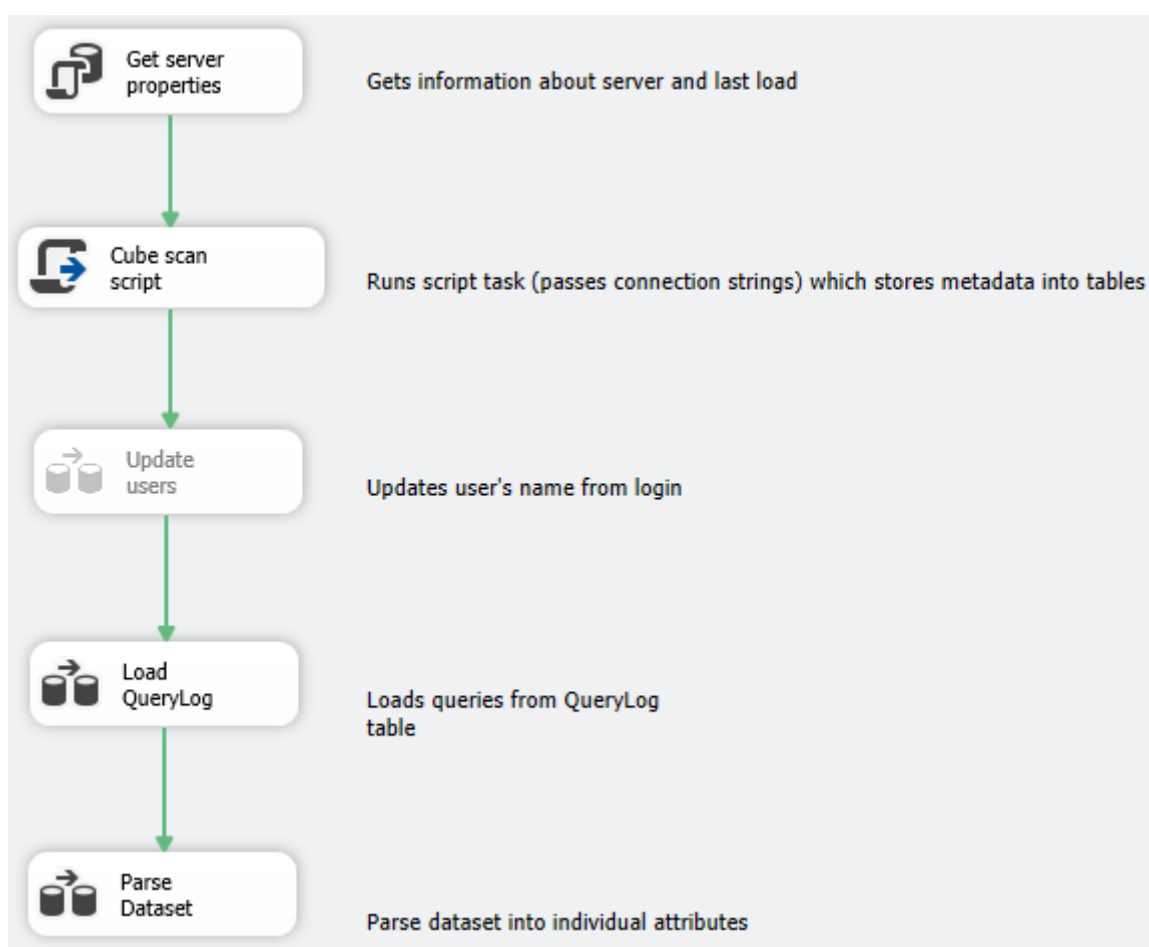
directory a pri načítaní reportov je navyše jeden “Lookup” task pre priradenie skupiny novým reportom.

Po skončení updatovania dimenzií nasleduje update uskutočnených zápisov kde sa zapíšu nové metadáta o daných zápisoch.

Nakoniec sa uskutoční zresetovanie parametrov a zápis do tabuľky prehľadu o tom koľko riadkov sa v jednotlivých updatovacích úlohach zapísalo. Ide o úlohu spustenia daného SQL kódu, ktorý obsahuje parametre, ktoré udávajú počet zapísaných riadkov. Spomínané vyresetovanie parametrov slúži na vymazanie údajov zo stĺpca pre parametre, s ktorými sa daný report spúšťa. To firmu v tomto momente nezaujíma avšak niekedy v budúcnosti môže nastať potreba tieto informácie taktiež evidovať.

2.3.3 Load Query Log

Druhý ETL process, ktorý má na starosti načítanie základných dát z používania OLAP kocky. Slúži na zápis dát, ktoré zobrazujú vyťaženosť OLAP kociek z analytického servera.



Obrázok č.13.: ETL pre zápis využívania kocky, Zdroj: Vlastná tvorba

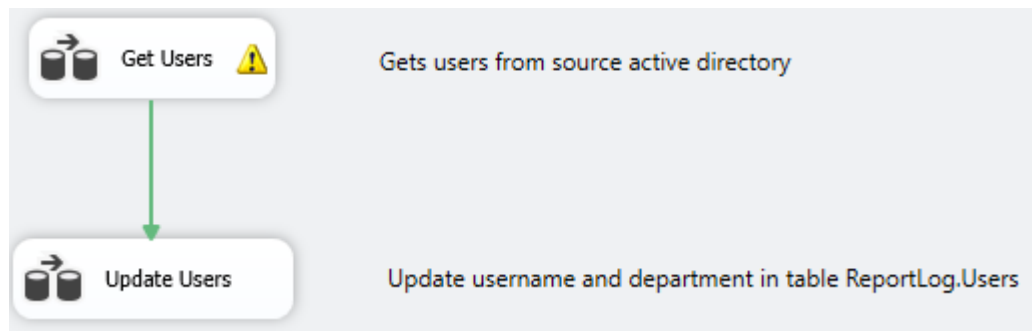
V tomto blíčku má prvý data-flow task na starosti najskôr získať informácie o serveroch a ich posledných načítaní a získať reťazce s napojením na ne. V ďalšom kroku je task, ktorý updatuje tabuľku užívateľov pomocou úlohy pomaly meniacej sa

dimenzie (Slowly Changing Dimension). Jedná sa o ten istý VB skript, ktorý je použitý v pri načítaní užívateľov v balíčku “OlapReportLog”, a zapisuje meno a deptment do tabuľky užívateľov. V tomto prípade je však tento task “vypnutý” pretože sa v projekte pristupuje k dátam užívateľov inou cestou a to cez OLAP active directory a na to bol vytvorený samostatný balíček “Search for users” popísaný neskôr.

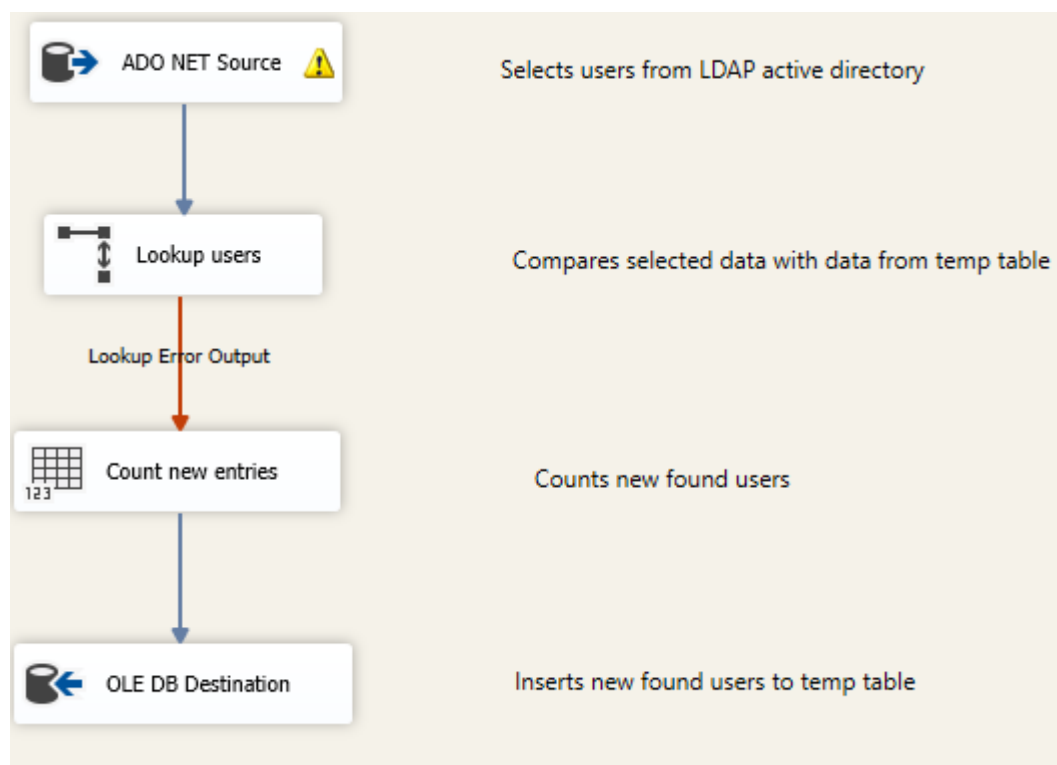
Posledné dva kroky sú načítanie jednotlivých záznamov podľa toho aký majú štartovací čas a rozdelenie jednotlivých atribútov do tabuliek podľa toho kam patria. Toto je spravené ďalším visual basic skriptom, pretože je potreba aby daný task prešiel každú dimenziu a vytvoril pre ňu vlastný SQL insert statement, čiže kód pre vloženie dát.

2.3.4 Search for users

Tento balíček má na starosti to aby do tabuľky užívateľov dotiahol meno užívateľa a department na ktorom pracuje. Balíček bol vytvorený pretože na danom serveri sa nachádzajú len prihlasovacie údaje o užívateľoch a nič iné. Naskôr je však potrebné spustiť balíček “Load Report Log” v ktorom sa načítajú užívatelia, ktorí majú prístup na server. Následne prvý data-flow task z balíčku dohľadávania užívateľov vytiahne dáta z OLAP active directory a zapíše ich do tabuľky “OLAP Active Directory. Zároveň tento data-flow task obsahuje aj kontrolu nových záznamov, ktoré v tabuľke ešte nie sú a tak zapíše len nové (unikátne) výskyty aby sa zbytočne predišlo duplicitám a v podstate aby sa skrátil čas pre vykonanie tejto ulohy na minimum čo je pri rádovo tisíckach riadkov zanedbateľné ale pri väčších objemoch dát to už môže byť pociťované.



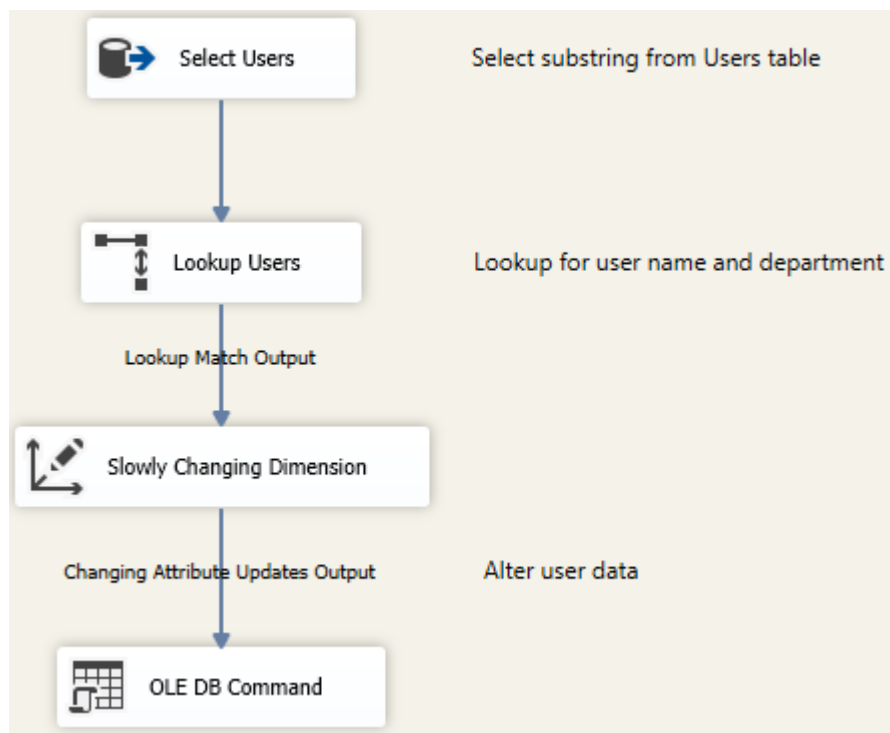
Obrázok č.14.: ETL pre získanie užívateľov z Active Directory a ich zápis do tabuľky užívateľov, Zdroj: Vlastná tvorba



Obrázok č.15.: Načítanie užívateľov z AD, Zdroj: Vlastná tvorba

Úlohou druhého data-flow tasku je to čo malo byť zahrnuté v balíčku “Load Query Log” pri update tabuľky užívateľov. Rozdelené do samostatného balíčka je to čisto z dôvodu prehľadnosti a spravovania. To znamená, že táto úloha najskôr vybere

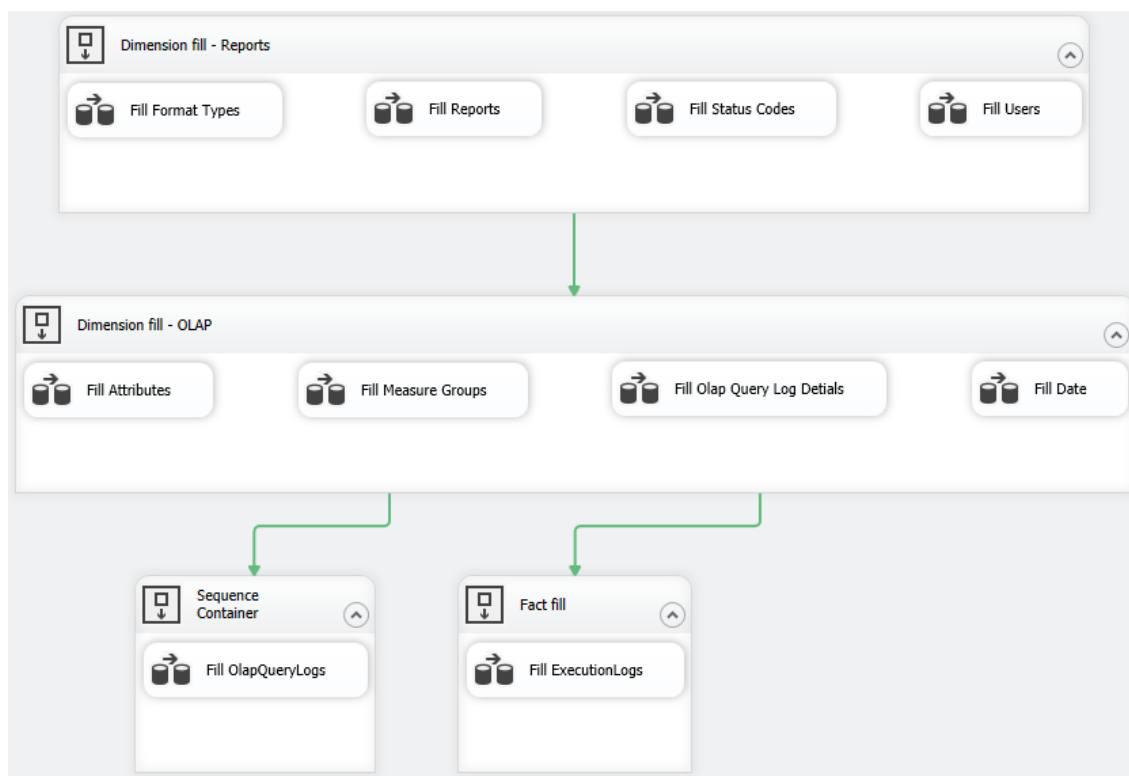
podreťazec atribútu UserLogin z tabuľky užívateľov orezaný o doménu, a potom ho porovná s atribútom z tabuľky záznamov z active directory a pomocou pomaly sa meniacej dimenzie upraví záznamy v tabuľke užívateľov tak, že doplní meno a department.



Obrázok č.16.: Úpdate pre tabuľku užívateľov, Zdroj: Vlastná tvorba

2.3.5 Vytvorenie tabuliek faktov a dimenzií

Nasledujúcim krokom po ukončení ETL procesu pre zber dát je ich očistenie o nepotrebné údaje a spracovanie pomocou ďalšieho ETL z relačnej databázy do dátového skladu a potom do multidimenzionálnej databázy čiže do kocky pre účel archivácie.



Obrázok č.17.: ETL pre naplnenie tabuliek dimenzií a tabuliek faktov

Najskôr však treba vytvoriť tabuľky faktov a tabuľky dimenzií. To znamená vytvorenie tabuliek s meranými veličinami a tabuliek s popisom daných veličín. Takto vytvorené tabuľky v dátovom sklade sa následne naplnia pomocou ďalšieho ETL balíčka postupne najskôr tabuľky dimenzií a následne tabuľky faktov. Dôležité je dodržať postupnosť plnenia, pretože pri plnení faktových tabuliek je nutnosť aby správne odkazovali na tabuľky dimenzií. Pri plnení tabuliek faktov je použitý “Lookup” proces na väzby, čiže dohľadávanie správnych väzieb voči tabuľkám dimenzií.

Ako základ pre toto riešenie je potreba vytvoriť osem tabuliek dimenzií a dve tabuľky faktov, čiže osem tabuliek s popisom a k nim dve tabuľky s meranými veličinami. Základné tabuľky dimenzií sú :

- **Tabuľka užívateľov**

Tabuľka užívateľov obsahuje všetkých užívateľov, ktorí sa nachádzajú v zložke active directory na serveri. Samotné údaje z active directory nevie data flow úloha priradiť k jednotlivým záznamom v logu spracovaní a tak je potreba ich naviazať cez tabuľku z Olap Query Log, kde sú uvedené ako login jednotlivých užívateľov.

- **Tabuľka reportov**

Tabuľka reportov obsahuje všetky reporty, ktoré sa nachádzajú v katalógu reportov na serveri. Je v nej zapísaný ako názov report, tak aj fyzická cesta k reportu, typ reportu a iné atribúty.

- **Tabuľka formátov**

Táto tabuľka obsahuje formáty, v ktorých je možné si jednotlivé reporty zobrazovať. V podstate ide o tabuľku podobnú číselníkovej tabuľke v relačnej databáze.

- **Tabuľka status kódov**

Tak isto ako tabuľka formátov sa jedná o číselníkovú tabuľku obsahujúcu status kódy určujúce či bol dotaz na dáta úspešný alebo nie.

- **Tabuľka atribútov**

Táto tabuľka zobrazuje atribúty, ktoré sa nachádzajú v multidimenzionálnej databáze, nad ktorou sa dané riešenie spúšťa. Ide o atribúty nachádzajúce sa v tabuľkách dimenzií v danej OLAP databáze.

- **Tabuľka measure group**

Táto tabuľka obsahuje informácie o jednotlivých meraných veličinách, ktoré sa nachádzajú vo faktových tabuľkách monitorovanej OLAP kocky. Jej obsah tvorí meno databázy, kde sa kocka nachádza, názov kocky, názov meranej veličiny a cesta v hierarchii jednotlivých veličín.

- **Tabuľka Olap Query Log Details**

Tabuľka určená na rozdelenie väzby many-to-many medzi tabuľkou OlapQueryLog a tabuľkou atribútov. Obsahuje len primárne kľúče z každej tabuľky.

- **Tabuľka dátumov**

Dátumová tabuľka určená obsahujúca kedy si jednotliví užívatelia prehliadali kocku a taktiež pre vytvorenie časových snímok na kocke, ktorá bola novovytvorená pre dáta z monitorovania.

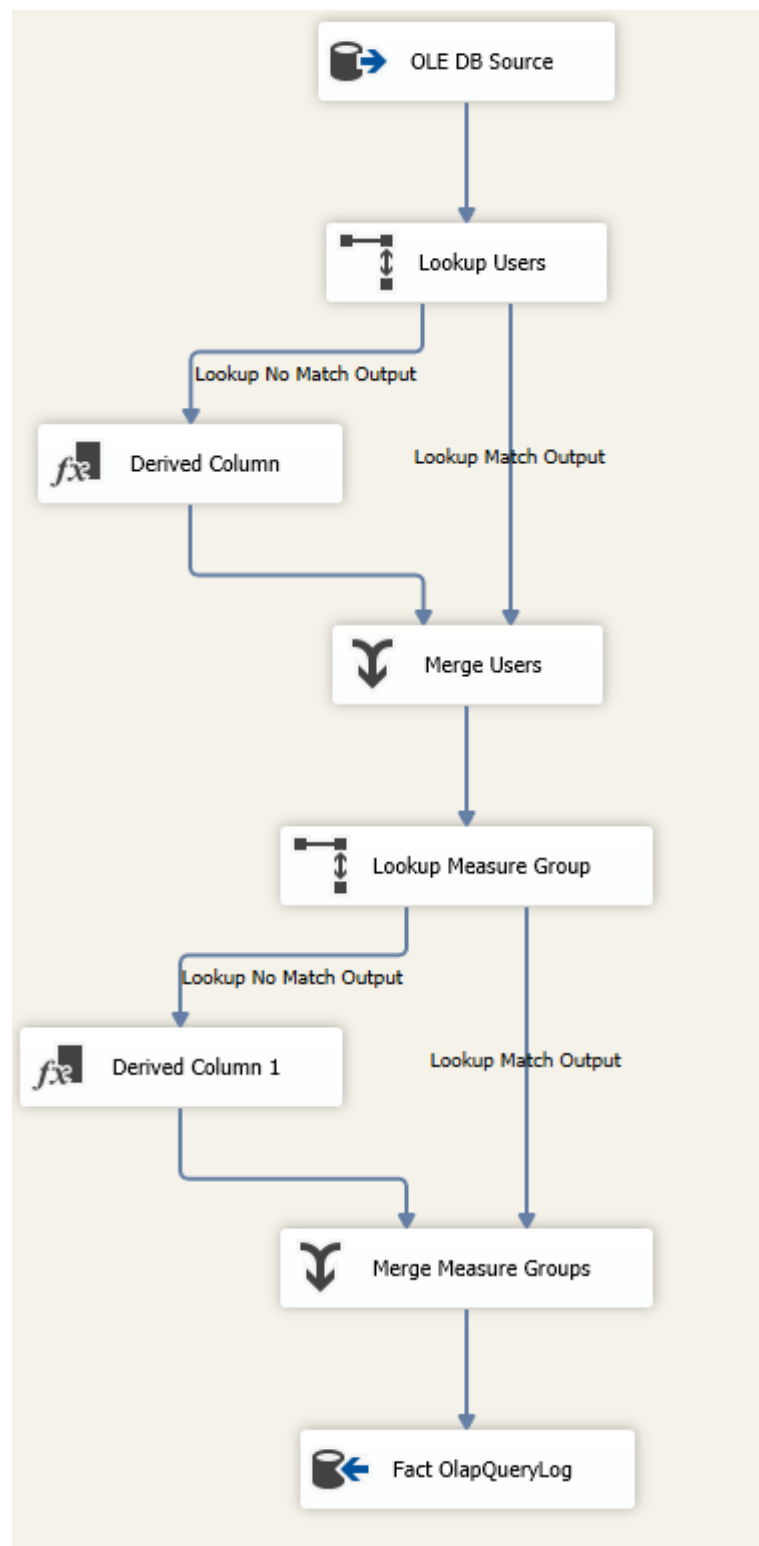
Tabuľky faktov sú dve a to:

- **Tabuľka Olap Query Logs**

Táto tabuľka obsahuje dátata, ktoré vznikajú pri prehliadaní kocky a to mimo iné hlavne počet načítaných riadkov a doba, ktorú zabralo samotné načítavanie dát v milisekundách.

- **Tabuľka Execution Logs**

Táto tabuľka obsahuje záznamy o prehliadaní, resp. spúšťaní repotov. Taktiež obsahuje okrem iného aj počty načítaných riadkov, dobu ako dlho trvalo načítanie, dobu vykresľovania reportu a aký veľký objem dát bol načítaný.



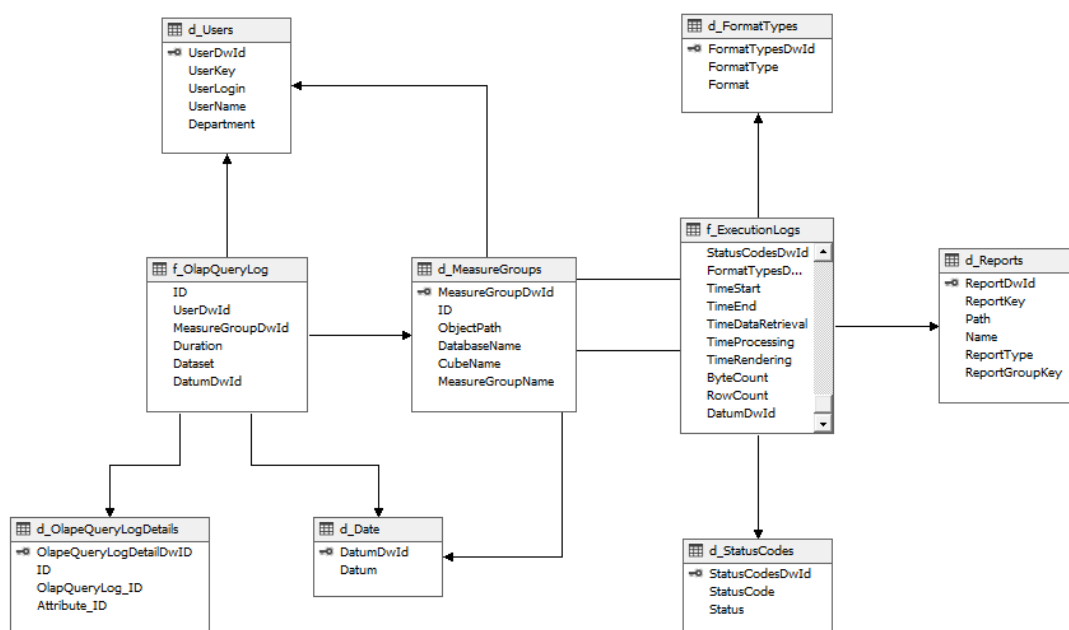
Obrázok č.18.: Data-flow task – detail plnenia faktovej tabuľky, Zdroj:

Vlastná tvorba

2.3.6 Vytvorenie multidimenzionálnej databázy

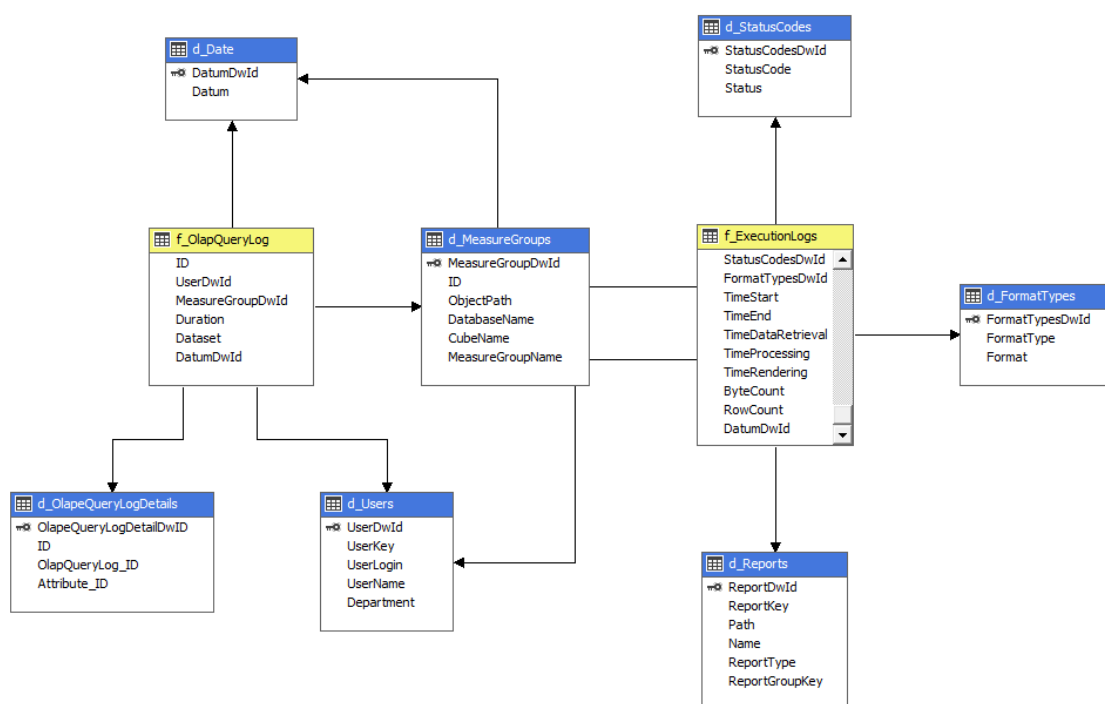
Posledným krokom v spracovávaní dát je vytvorenie multidimenzionálnej databázy a jej naplnenie vytvorenými tabuľkami faktov a dimenzií. V takto vytvorenej multidimenzionálnej databáze sa budú ukladať historické dáta z predchádzajúcich krokov a následne sa z týchto dát budú vytvárať reporty a to v prostredí Power BI alebo pomocou MDX dotazov v prostredí reportovacích služieb.

Postup pri vytváraní dátovej kocky začína vytvorením nového multidimenzionálneho projektu v programe Visual Studio. Postupne si treba vytvoriť dátový zdroj, z ktorého sa čerpajú dáta do multidimenzionálnej databázy a na to vytvoriť pohľad, ktorý zobrazuje väzby medzi tabuľkami.



Obrázok č.19.: Pohľad na dátový zdroj pri tvorbe multidimenzionálnej databázy, Zdroj: Vlastná tvorba

Následne na to je nutné určiť, ktoré tabuľky sú dimenzie, a ktoré sú faktové tabuľky. Pri tomto sa vyberajú primárne kľúče, pomocou ktorých sa jednotlivé dáta v multidimenzionálnej databáze budú dohľadávať, a tak isto aj atribúty, ktoré chceme aby sa v kocke zobrazovali. Na koniec tohoto procesu je potrebné vytvoriť dátovú kocku a potom ju nasadiť na určený analytický server a nechať spracovať. Po tomto je možné si takto vytvorenú kocku prehliadať ako vo Visual Studiu tak aj v Management Studiu.



Obrázok č.20.: Multidimenzionálna databáza – kocka, Zdroj: Vlastná tvorba

Na obrázku je vidieť ako približne vyzerá multidimenzionálna databáza. Samozrejmosťou je, že sa projekt bude postupne realizovať naďalej. V tom prípade je možné, že sa niektoré tabuľky dimenzií ukážu ako nepotrebné alebo nejaké nové tabuľky do kocky pribudnú. To sa však môže líšiť podľa toho na aký projekt sa toto riešenie nasadí. V podstate je ale vidieť základ multidimenzionálnej databázy

2.3.7 Reporty

Po skončení časti s dátovými tokmi je potreba tieto dáta zobrazit' tak aby sa z nich dalo prehľadne čítať. Ide teda o vytvorenie reportov či už v tabuľkovej alebo grafickej podobe. Jedná sa o administratívne reporty, ktoré budú slúžiť len pre interné účely firmy a k jednotlivým zákazníkom sa nedostanú.



Obrázok č.21.: Microsoft Visual Studio, Zroj: images.google.com

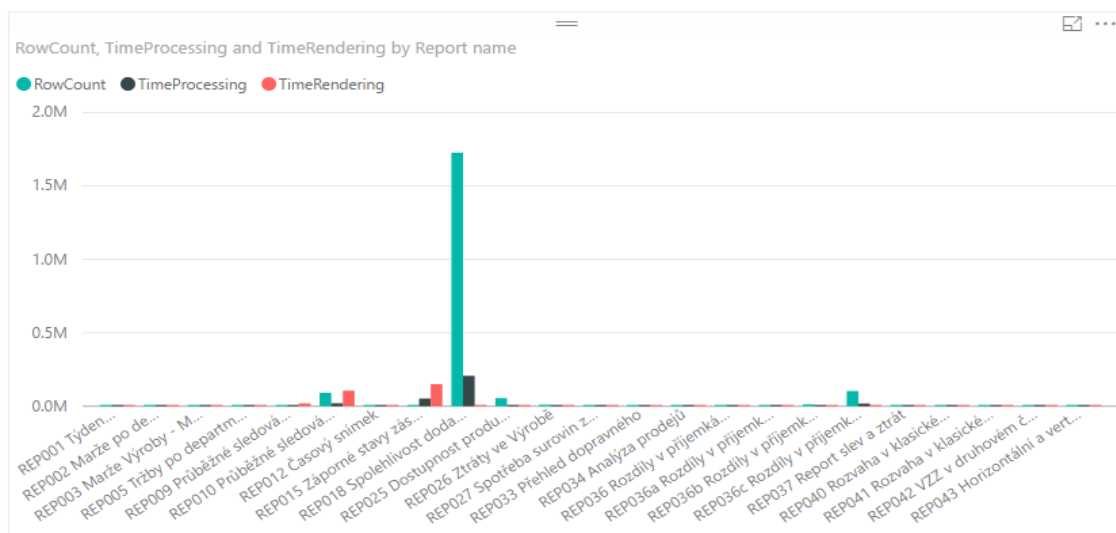
Hlavnými reportami, ktorými sa táto práca zaoberá sú dva. Jeden bude report pre analýzu využitia kociek a to použitím MDX dotazov alebo iných prostriedkov, napríklad pomocou zobrazenia dát v exceli popřípade v programe Visual Studio pomocou SSRS a druhý report bude zameraný na sprehladnenie využívania a zobrazovania jednotlivých reportov a bude spracovaný a zobrazený v programe Power BI.



Obrázok č.22.: Power BI, Zdroj: images.google.com

2.3.8 Zobrazenie prehliadania reportov

Pre vizualizáciu toho ako sú jednotlivé reporty využívané je použité prostredie programu Power BI kde je prehľadne vyzobrazené ako moc sa jednotlivé reporty využívajú, kto ich využíva, ako dlho trvá než sa načítajú dáta alebo aj to ako dlho trvá kým sa dáta v reportoch zobrazia.



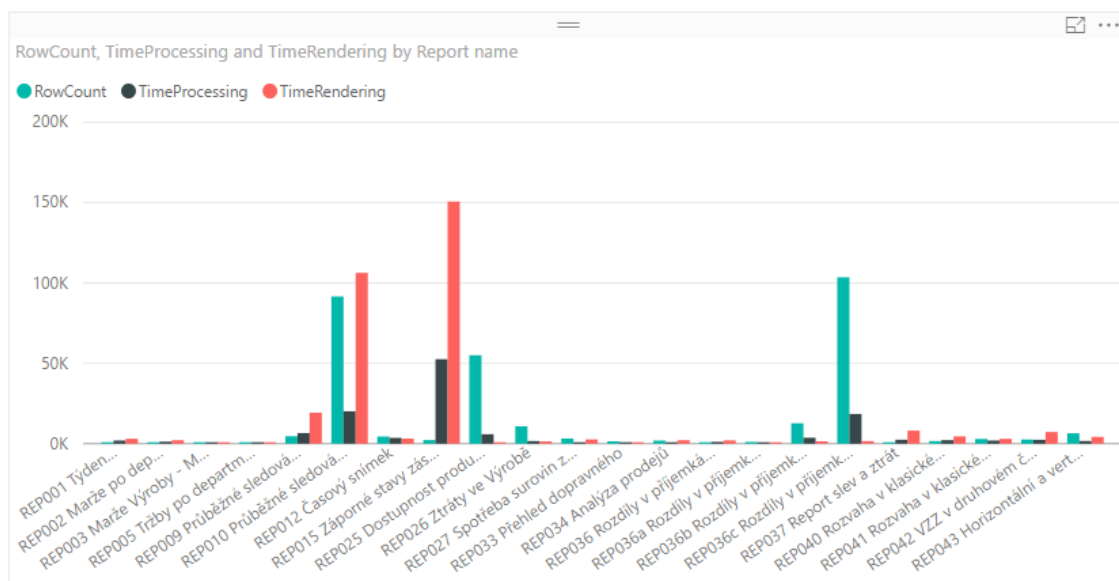
Obrázok č.23.: Zobrazenie využívanie reportov v Power Bi, Zdroj: Vlastná tvorba

Na prvom obrázku je vidieť, že najväčší počet zobrazených riadkov, ktoré sa počítajú pri spustení daného report a následne zapíšu do logu, má report č.18 zobrazujúci spoľahlivosť. V porovnaní s ostatnými reportami je najviditeľnejší preto, lebo samotný zber dát sa začal neskôr ako boli vytvorené ostatné reporty avšak report č.18 bol vytvorený neskôr a následne bol aj upravovaný.

| Name | UserName | RowCount | TimeStart |
|-------------------------|----------------|----------------|----------------------|
| REP018 Spolehlivost ... | Jakub Kmet' | 131943 | 5/22/2017 3:13:07 PM |
| REP018 Spolehlivost ... | Jiří Klimeš | 66591 | 3/29/2017 1:53:34 PM |
| REP018 Spolehlivost ... | Jiří Klimeš | 66591 | 3/31/2017 4:48:02 PM |
| REP018 Spolehlivost ... | Jiří Klimeš | 4734 | 3/31/2017 4:53:16 PM |
| REP018 Spolehlivost ... | Jiří Klimeš | 66591 | 3/31/2017 4:55:32 PM |
| REP018 Spolehlivost ... | Jiří Klimeš | 128959 | 4/6/2017 1:53:11 PM |
| REP018 Spolehlivost ... | Jiří Klimeš | 134467 | 4/6/2017 1:57:12 PM |
| REP018 Spolehlivost ... | Jiří Klimeš | 131943 | 5/19/2017 3:27:11 PM |
| REP018 Spolehlivost ... | Jiří Klimeš | 152607 | 5/19/2017 3:28:08 PM |
| REP018 Spolehlivost ... | Ondřej Havelka | 66591 | 3/27/2017 3:18:47 PM |
| REP018 Spolehlivost ... | Ondřej Havelka | 66591 | 3/27/2017 3:20:53 PM |
| REP018 Spolehlivost ... | Ondřej Havelka | 0 | 3/27/2017 3:21:51 PM |
| REP018 Spolehlivost ... | Ondřej Havelka | 66871 | 4/6/2017 10:48:56 AM |
| REP018 Spolehlivost ... | Ondřej Havelka | 128959 | 4/6/2017 12:15:37 PM |
| REP018 Spolehlivost ... | Ondřej Havelka | 127231 | 5/11/2017 1:46:52 PM |
| REP018 Spolehlivost ... | Ondřej Havelka | 127231 | 5/11/2017 2:25:10 PM |
| Total | | 1726433 | |

Obrázok č.24.: Zobrazenie využívanie reportov v Power Bi, Zdroj: Vlastná tvorba

Samozrejmou je aj to, že jednotlivé reporty nezobrazujú rovnaké počty riadkov, a preto tam sú viditeľné rozdiely.



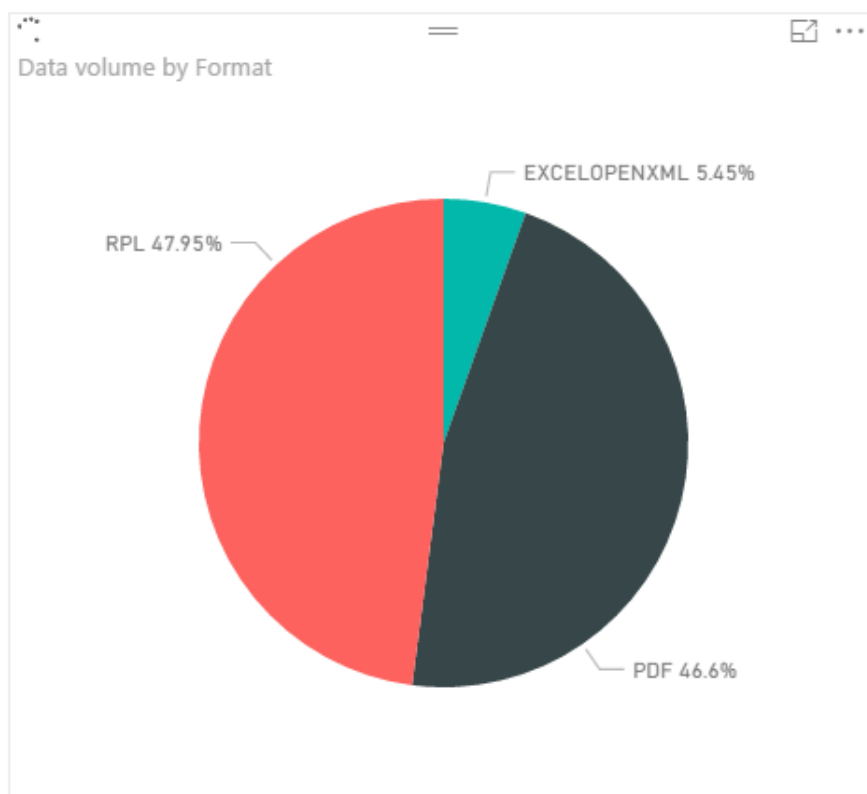
Obrázok č.25.: Zobrazenie využívanie reportov v Power Bi, Zdroj: Vlastná tvorba

Na tomto obrázku je možno vidieť to čo na prvom obrázku z Power BI, avšak tu bol z filtra odstránený report č.18 a ostatné reporty tam ostali. Tu už lepšie vidieť rozdiely medzi ostatnými reportami, pretože nie sú zmenšené kvôli danému report č.18.

| Name | UserName | RowCount | TimeStart |
|--------------|-----------------|-------------|----------------------|
| REP015 ... | DWH SQL Service | 9 | 5/13/2017 1:12:22 AM |
| REP015 ... | DWH SQL Service | 0 | 5/13/2017 1:12:27 AM |
| REP015 ... | DWH SQL Service | 3 | 5/13/2017 1:12:29 AM |
| REP015 ... | DWH SQL Service | 21 | 5/13/2017 1:12:31 AM |
| REP015 ... | DWH SQL Service | 5 | 5/13/2017 1:12:33 AM |
| REP015 ... | DWH SQL Service | 4 | 5/14/2017 1:15:42 AM |
| REP015 ... | DWH SQL Service | 52 | 5/14/2017 1:15:48 AM |
| REP015 ... | DWH SQL Service | 2 | 5/14/2017 1:15:50 AM |
| REP015 ... | DWH SQL Service | 1 | 5/14/2017 1:15:52 AM |
| REP015 ... | DWH SQL Service | 5 | 5/14/2017 1:15:54 AM |
| REP015 ... | DWH SQL Service | 11 | 5/19/2017 1:13:34 AM |
| REP015 ... | DWH SQL Service | 3 | 5/19/2017 1:13:41 AM |
| REP015 ... | DWH SQL Service | 2 | 5/19/2017 1:13:43 AM |
| REP015 ... | DWH SQL Service | 21 | 5/19/2017 1:13:45 AM |
| REP015 ... | DWH SQL Service | 2 | 5/19/2017 1:13:48 AM |
| REP015 ... | DWH SQL Service | 8 | 5/20/2017 1:17:09 AM |
| REP015 ... | DWH SQL Service | 4 | 5/20/2017 1:17:15 AM |
| REP015 ... | DWH SQL Service | 3 | 5/20/2017 1:17:18 AM |
| REP015 ... | DWH SQL Service | 19 | 5/20/2017 1:17:20 AM |
| REP015 ... | DWH SQL Service | 3 | 5/20/2017 1:17:22 AM |
| REP015 ... | DWH SQL Service | 13 | 5/22/2017 2:31:29 AM |
| REP015 ... | DWH SQL Service | 1 | 5/22/2017 2:31:35 AM |
| REP015 ... | DWH SQL Service | 1 | 5/22/2017 2:31:37 AM |
| REP015 ... | DWH SQL Service | 13 | 5/22/2017 2:31:39 AM |
| REP015 ... | DWH SQL Service | 2 | 5/22/2017 2:31:41 AM |
| REP015 ... | Jiří Klimeš | 47 | 4/6/2017 2:39:21 PM |
| REP015 ... | Jiří Klimeš | 47 | 4/6/2017 3:31:03 PM |
| REP015 ... | Jiří Klimeš | 7 | 4/6/2017 3:31:18 PM |
| REP015 ... | Jiří Klimeš | 56 | 4/7/2017 4:51:05 PM |
| REP015 ... | Jiří Klimeš | 74 | 4/7/2017 4:51:13 PM |
| REP015 ... | Jiří Klimeš | 3 | 4/7/2017 4:51:26 PM |
| REP015 ... | Luděk Šišák | 80 | 4/4/2017 11:30:47 AM |
| REP015 ... | Luděk Šišák | 6 | 4/4/2017 11:30:55 AM |
| REP015 ... | Luděk Šišák | 80 | 4/4/2017 5:08:40 PM |
| REP015 ... | Luděk Šišák | 7 | 4/4/2017 5:08:52 PM |
| REP015 ... | Luděk Šišák | 62 | 4/24/2017 4:29:08 PM |
| REP015 ... | Luděk Šišák | 30 | 4/24/2017 4:29:33 PM |
| REP015 ... | Luděk Šišák | 40 | 4/24/2017 4:29:53 PM |
| Total | | 2255 | |


Obrázok č.26.: Zobrazenie využívanie reportov v Power Bi, Zdroj: Vlastná tvorba

Zaujímavosťou je napríklad report č.15 kde je vidieť, že renderovací čas reportu je omnoho vyšší v porovnaní s počtom vyzobrazených riadkov. To má za následok automatická SQL služba, ktorá každý deň nadržanom zasiela report za predchádzajúci deň.



Obrázok č.27.: Podiel formátov v akých boli reporty zobrazené, Zdroj: Valstná tvorba

Na tomto obrázku je vidieť množstvo dát v závislosti na formáte v akom sa jednotlivé reporty zobrazujú. Najväčší objem dát, skoro až 48 percent zaberá zobrazovanie cez webový prehliadač z reportovacieho protálu, v tesnom závесе za ním je zobrazovanie v PDF formáte so skoro 47 percentami a najmenej objemu dát zaberá prehliadanie reportov vo forme excelu.

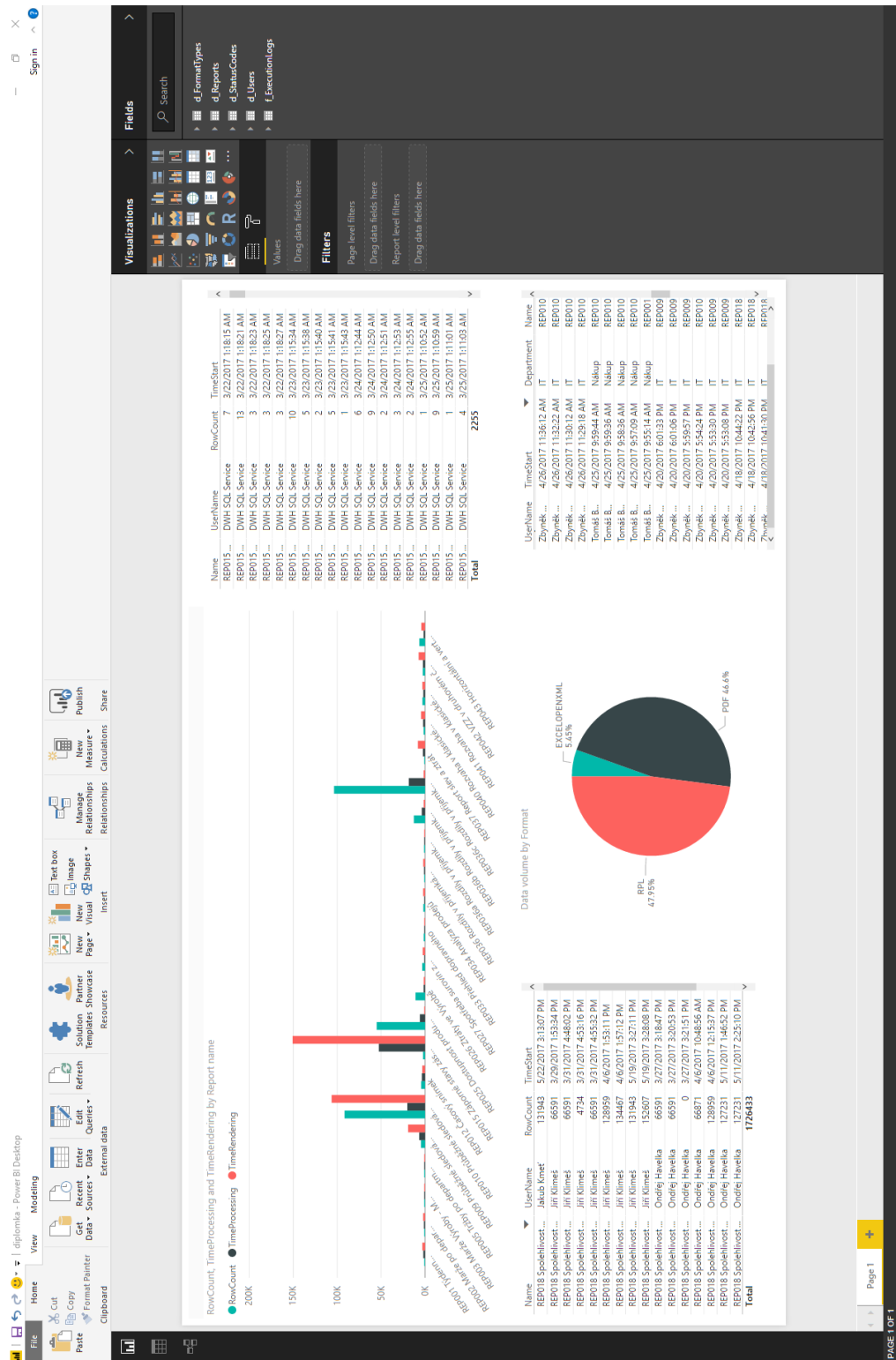


| UserName | TimeStart | Department | Name |
|------------|-----------------------|------------|--------|
| Tomáš B... | 4/28/2017 10:40:09 AM | Nákup | REP010 |
| Tomáš B... | 4/28/2017 10:39:50 AM | Nákup | REP010 |
| Zbyněk ... | 4/26/2017 11:36:21 AM | IT | REP010 |
| Zbyněk ... | 4/26/2017 11:36:12 AM | IT | REP010 |
| Zbyněk ... | 4/26/2017 11:32:22 AM | IT | REP010 |
| Zbyněk ... | 4/26/2017 11:30:12 AM | IT | REP010 |
| Zbyněk ... | 4/26/2017 11:29:18 AM | IT | REP010 |
| Tomáš B... | 4/25/2017 9:59:44 AM | Nákup | REP010 |
| Tomáš B... | 4/25/2017 9:59:36 AM | Nákup | REP010 |
| Tomáš B... | 4/25/2017 9:58:36 AM | Nákup | REP010 |
| Tomáš B... | 4/25/2017 9:57:09 AM | Nákup | REP010 |
| Tomáš B... | 4/25/2017 9:55:14 AM | Nákup | REP001 |
| Zbyněk ... | 4/20/2017 6:01:33 PM | IT | REP009 |
| Zbyněk ... | 4/20/2017 6:01:06 PM | IT | REP009 |
| Zbyněk ... | 4/20/2017 5:59:57 PM | IT | REP009 |
| Zbyněk ... | 4/20/2017 5:54:24 PM | IT | REP010 |
| Zbyněk ... | 4/20/2017 5:53:30 PM | IT | REP009 |
| Zbyněk ... | 4/20/2017 5:53:08 PM | IT | REP009 |
| Zbyněk ... | 4/18/2017 10:44:22 PM | IT | REP018 |
| Zbyněk ... | 4/18/2017 10:42:56 PM | IT | REP018 |
| Zbyněk ... | 4/18/2017 10:41:30 PM | IT | REP018 |
| Zbyněk ... | 4/18/2017 7:33:50 PM | IT | REP018 |
| Radek S... | 4/13/2017 2:39:39 PM | Logistika | REP010 |
| Radek S... | 4/13/2017 2:39:27 PM | Logistika | REP010 |
| Radek S... | 4/13/2017 2:38:59 PM | Logistika | REP010 |

Obrázok č.28.: Zobrazenie užívateľov a reportov, ktoré si pozerali, Zdroj:

Vlastná tvorba

Tento report zobrazuje kto a z akého oddelenia si prehliada aké reporty. Toto by sa dalo využiť pri automatizovaní reportov a ich zasielania. Napríklad pán Tomáš si prezerá reporty č.10 a č.1 v približne rovnaký čas pred desiatou hodinou ráno. To znamená, že by bolo možné nastaviť automatické spracovanie a zaslanie reportov na čas, kedy by bol server menej vyťažovaný. Tak by pán Tomáš mal pripravený report č.10 v jeho e-mailovej schránke a nemusel by si ho sám dohľadávať. Takýmto spôsobom by sa mohli uvoľniť zdroje a výpočtový výkon servera cez deň, kedy sa zobrazujú reporty, ktoré si užívatelia zapínajú nepravidelne.



Obrázok č.29.: Prostredie PowerBI, Zdroj: Vlastná tvorba

2.3.9 Zobrazenie vyt'ažnosti kocky

Reporty vyt'aženia kocky zobrazujú to ako užívatelia pristupujú k dátam a aké hodnoty ich vlastne zaujímajú. Formou pripomínajú tabuľky, ktoré sa dajú vytvoriť priamo v exceli avšak prostredie programu Visual studio ponúka iné možnosti ako Excel ako taký. V exceli si užívatelia môžu prehliadať kocku ako cez Management Studio, avšak oproti nemu môžu vytvorené tabuľky ľubovoľne formátovať čo v Management Studio nie je možné. Narozdiel od Excelu slúži Visual Studio na vytváranie reportov v danej forme, to znamená, že sa vytvorí report podľa toho čo chce mať zákazník zobrazené. Takto nadefinované reporty je možné nastaviť napríklad na automatické vytvorenie a odoslanie užívateľom a to buď v podobe Excelovského zošitu alebo v podobe PDF a to všetko v prednastavený čas.

| User Name | Datum | Duration in miliseconds | Number of rows |
|------------------------|------------|-------------------------|----------------|
| DWH Reporting Services | 2017-05-11 | 8120 | 879 |
| DWH Reporting Services | 2017-05-12 | 84 | 156 |
| DWH Reporting Services | 2017-05-13 | 561 | 168 |
| DWH Reporting Services | 2017-05-18 | 108 | 186 |
| Luděk Šišák | 2017-05-12 | 435 | 456 |
| Luděk Šišák | 2017-05-15 | 3762 | 48 |
| Luděk Šišák | 2017-05-16 | 6 | 54 |
| Luděk Šišák | 2017-05-19 | 111 | 87 |
| Ondřej Havelka | 2017-05-05 | 155 | 1 |
| Ondřej Havelka | 2017-05-11 | 5047 | 84 |
| Pavel Urban | 2017-05-15 | 3558 | 879 |
| Pavel Urban | 2017-05-19 | 80 | 60 |
| Radek Soukup | 2017-05-11 | 2088 | 800 |
| Radek Soukup | 2017-05-12 | 84 | 204 |

Obrázok č.30.: Report z prostredia Visual Studio, Zdroj: Vlastná tvorba

V tomto reporte sú vidieť dni, v ktoré k nim užívatelia pristupujú, ako dlho trvá než sa im dáta načítajú a počty riadkov ku ktorým sa daný užívateľ dopracoval.

| User Name | Cube Name | Measure Group Name | Number of rows | Duration in milliseconds |
|------------------------|-----------|--------------------|----------------|--------------------------|
| DWH Reporting Services | Finance | Rozvaha | 10 | 5 |
| DWH Reporting Services | Finance | VZZ | 39 | 44 |
| DWH Reporting Services | Prodeje | Dodávky | 28 | 104 |
| DWH Reporting Services | Prodeje | Footfall | 30 | 18 |
| DWH Reporting Services | Prodeje | Odpisy | 70 | 175 |
| DWH Reporting Services | Prodeje | Počet prodejů | 36 | 588 |
| DWH Reporting Services | Prodeje | Prodeje | 191 | 5319 |
| DWH Reporting Services | Prodeje | Příjemky | 40 | 280 |
| DWH Reporting Services | Prodeje | Slevy | 35 | 230 |
| DWH Reporting Services | Prodeje | Spotřeba | 35 | 90 |
| DWH Reporting Services | Prodeje | Výdejky | 77 | 1111 |
| DWH Reporting Services | Prodeje | Zásoby | 733 | 868 |
| DWH Reporting Services | Prodeje | Zásoby záporné | 65 | 41 |
| Luděk Šišák | Prodeje | Odpisy | 21 | 6 |
| Luděk Šišák | Prodeje | Prodeje | 576 | 4287 |
| Luděk Šišák | Prodeje | Spotřeba | 48 | 21 |
| Ondřej Havelka | Prodeje | Prodeje | 63 | 4984 |
| Ondřej Havelka | Prodeje | Výroba | 21 | 63 |
| Ondřej Havelka | Unknown | Unknown | 1 | 155 |
| Pavel Urban | Finance | VZZ | 21 | 531 |
| Pavel Urban | Prodeje | Odpisy | 663 | 1735 |
| Pavel Urban | Prodeje | Prodeje | 48 | 201 |
| Pavel Urban | Prodeje | Spotřeba | 207 | 1171 |
| Radek Soukup | Prodeje | Odpisy | 616 | 196 |
| Radek Soukup | Prodeje | Prodeje | 216 | 1696 |
| Radek Soukup | Prodeje | Zásoby | 172 | 280 |

Obrázok č.31.: Report prehliadania OLAP servera, Zdroj: Vlastná tvorba

V tomto reporte je zas vidieť, ktoré kocky si užívatelia prehliadajú, ktoré hodnoty si nechávajú zobrazit', a taktiež koľko riadkov im bol zobrazené a ako dlho to trvalo v milisekundách.

Takýchto záznamov je v porovnaní so záznamami o prehliadaní reportov omnoho menej, pretože užívatelia sú málokedy nútení ísť prehliadať kocku priamo cez management studio a práve na to slúžia samotné reporty. Avšak niekedy treba pristúpiť k dátam iným spôsobom ako cez reporty. To ale nie je až tak časté, a preto je lepšou voľbou ísť priamo ako mať na všetky možné kombinácie vytvorené reporty. To by znamenalo zbytočne veľké množstvo zabraných dát na serveri, ktoré by boli nevyužívané.

Report Project1 - Microsoft Visual Studio

File Edit View Project Build Debug Tools Report Format Default Debug Design Preview

Report Data

Report Explorer

Search Solution Explorer (Ctrl+J)

Solution Explorer

Report Project1 (1 project)

Shared Data Sources

Reports

Report prehladania kocky.rdl

Report prehladania kocky.rdl [Design]

Report prehladania kocky

| User Name | Datum | Duration in milliseconds | Number of rows |
|------------------------|------------|--------------------------|----------------|
| DWH Reporting Services | 2017-05-11 | 8120 | 879 |
| DWH Reporting Services | 2017-05-12 | 84 | 156 |
| DWH Reporting Services | 2017-05-13 | 561 | 168 |
| DWH Reporting Services | 2017-05-18 | 108 | 186 |
| Luděk Šišák | 2017-05-12 | 435 | 456 |
| Luděk Šišák | 2017-05-15 | 3762 | 48 |
| Luděk Šišák | 2017-05-16 | 6 | 54 |
| Luděk Šišák | 2017-05-19 | 111 | 87 |
| Ondřej Havella | 2017-05-05 | 155 | 1 |
| Ondřej Havella | 2017-05-11 | 5047 | 84 |
| Pavel Urban | 2017-05-15 | 3558 | 879 |
| Pavel Urban | 2017-05-19 | 80 | 60 |
| Radek Soukup | 2017-05-11 | 2088 | 800 |
| Radek Soukup | 2017-05-12 | 84 | 204 |

| User Name | Cube Name | Measure Group Name | Number of rows | Duration in milliseconds |
|------------------------|-----------|--------------------|----------------|--------------------------|
| DWH Reporting Services | Finance | Rozvaha | 10 | 5 |
| DWH Reporting Services | Finance | VZ | 39 | 44 |
| DWH Reporting Services | Prodej | Dodavky | 28 | 104 |
| DWH Reporting Services | Prodej | Football | 30 | 18 |
| DWH Reporting Services | Prodej | Odpisy | 70 | 175 |
| DWH Reporting Services | Prodej | Počet prodajů | 36 | 588 |
| DWH Reporting Services | Prodej | Příjemky | 191 | 5319 |
| DWH Reporting Services | Prodej | Slevy | 40 | 280 |
| DWH Reporting Services | Prodej | Spotřeba | 35 | 230 |
| DWH Reporting Services | Prodej | Výdělky | 35 | 90 |
| DWH Reporting Services | Prodej | Zásoby | 77 | 1111 |
| DWH Reporting Services | Prodej | Zapomně | 733 | 868 |
| DWH Reporting Services | Prodej | Odpisy | 65 | 41 |
| Luděk Šišák | Prodej | Odpisy | 21 | 6 |
| Luděk Šišák | Prodej | Prodej | 576 | 4287 |
| Luděk Šišák | Prodej | Spotřeba | 48 | 21 |
| Ondřej Havella | Prodej | Prodej | 63 | 4994 |
| Ondřej Havella | Prodej | Výroba | 21 | 63 |
| Ondřej Havella | Unknown | Unknown | 1 | 155 |
| Pavel Urban | Finance | VZ | 21 | 531 |
| Pavel Urban | Prodej | Odpisy | 663 | 1735 |
| Pavel Urban | Prodej | Prodej | 48 | 201 |
| Pavel Urban | Prodej | Spotřeba | 207 | 1171 |
| Radek Soukup | Prodej | Odpisy | 616 | 196 |
| Radek Soukup | Prodej | Prodej | 216 | 1696 |
| Radek Soukup | Prodej | Zásoby | 172 | 280 |

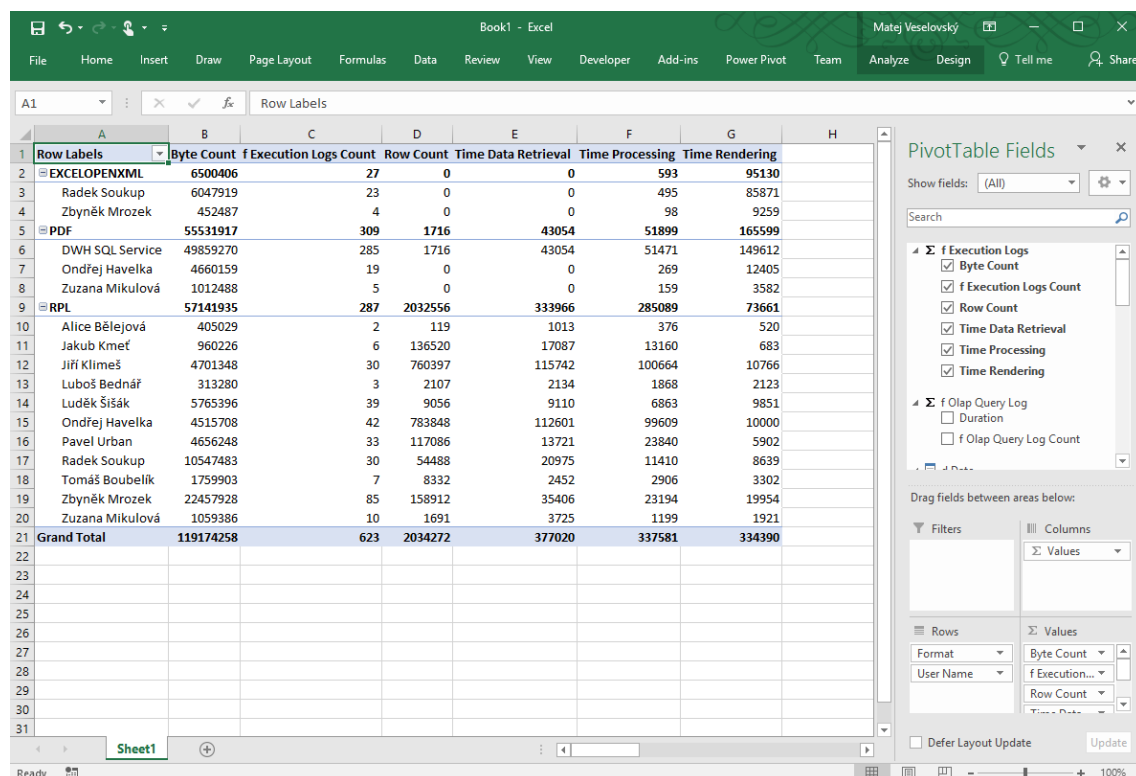
Report Data SSS Toolbox

Ready

Obrázok č.32.: Prostredie Visual Studio, Zdroj: Vlastná tvorba

2.3.10 Reporty v MS Excel

Užívatelia sa taktiež môžu na kocku pripojiť pomocou Excelu. To sa využíva hlavne vtedy, keď je potrebné vidieť ako dáta v kocke vyzerajú v rámci analytickej práce. Oproti webovým reportom je to jednoduchšia cesta ako sa dostať k dátam z kocky, avšak časovo o niečo náročnejšia a vyžaduje určitý level vzdelania resp. skúseností s prácou v exceli.



| Row Labels | Byte Count | f Execution Logs Count | Row Count | Time Data Retrieval | Time Processing | Time Rendering |
|-----------------|------------|------------------------|-----------|---------------------|-----------------|----------------|
| EXCELOPENXML | 6500406 | 27 | 0 | 0 | 593 | 95130 |
| Radek Soukup | 6047919 | 23 | 0 | 0 | 495 | 85871 |
| Zbyněk Mrozek | 452487 | 4 | 0 | 0 | 98 | 9259 |
| PDF | 55531917 | 309 | 1716 | 43054 | 51899 | 165599 |
| DWH SQL Service | 49859270 | 285 | 1716 | 43054 | 51471 | 149612 |
| Ondřej Havelka | 4660159 | 19 | 0 | 0 | 269 | 12405 |
| Zuzana Mikulová | 1012488 | 5 | 0 | 0 | 159 | 3582 |
| RPL | 57141935 | 287 | 2032556 | 333966 | 285089 | 73661 |
| Alice Bělejová | 405029 | 2 | 119 | 1013 | 376 | 520 |
| Jakub Kmeť | 960226 | 6 | 136520 | 17087 | 13160 | 683 |
| Jiří Klimeš | 4701348 | 30 | 760397 | 115742 | 100664 | 10766 |
| Luboš Bednář | 313280 | 3 | 2107 | 2134 | 1868 | 2123 |
| Luděk Šišák | 5765396 | 39 | 9056 | 9110 | 6863 | 9851 |
| Ondřej Havelka | 4515708 | 42 | 783848 | 112601 | 99609 | 10000 |
| Pavel Urban | 4656248 | 33 | 117086 | 13721 | 23840 | 5902 |
| Radek Soukup | 10547483 | 30 | 54488 | 20975 | 11410 | 8639 |
| Tomáš Boubelík | 1759903 | 7 | 8332 | 2452 | 2906 | 3302 |
| Zbyněk Mrozek | 22457928 | 85 | 158912 | 35406 | 23194 | 19954 |
| Zuzana Mikulová | 1059386 | 10 | 1691 | 3725 | 1199 | 1921 |
| Grand Total | 119174258 | 623 | 2034272 | 377020 | 337581 | 334390 |

Obrázok č.33.: Prehliadanie OLAP kocky v exceli, Zdroj: Vlastná tvorba

Zjednodušene povedané sú reporty na webovom portáli užívateľsky prívetivejšie a rýchlejšie a prehliadanie kocky v MS Excel slúži skôr na príležitostné prehliadanie kombinácií dát, pre ktoré nie sú vytvorené reporty.

2.3.11 Ďalšie spracovanie

Po dokončení testovacej fázy, ktorá bude trvať niekoľko mesiacov, sa finálna podoba riešenia bude dať nasadiť na akýkoľvek projekt, resp. dátový sklad. Jednotlivé ETL balíčky budú automatizované pomocou serverového agenta v MS SQL Server Management Studio. Reporty sa taktiež dajú nastaviť tak aby sa generovali automaticky a následne odosielali e-mailom tým užívateľom, ktorí ich budú potrebovať pri svojej práci rozdelené podľa jednotlivých projektov.

Záver

Cieľom tejto diplomovej práce bolo navrhnúť a vytvoriť monitorovacie a profilovacie riešenie nad BI systémom. Toto riešenie bolo navrhnuté podľa výsledku analýzy súčasného stavu a podľa potrieb firmy.

Vstupom do takto navrhnutého riešenia sú údaje zbierané ako z využívania OLAP servera tak aj jednotlivých reportov, ktoré sú uložené v katalógu reportov na serveri, následným spracovaním týchto dát pomocou ETL balíčkov nasjkôr do operačnej databázy, potom do dátového skladu a nakoniec do multidimenzionálnej databázy. Výstupom práce sú vzorové reporty vytvorené zo získaných dát a to jeden v prostredí Power BI a druhý v prostredí Visaul Studio 2015.

Jednotlivé dátové toky v ETL balíčkoch boli spracované pre čo najvyššiu prehľadnosť a držali sa zaužívaného štandardu pre ich vytváranie vo firme.

Aj keď je toto riešenie vo svojej testovacej fáze, už je vidieť aké dáta sa dajú pomocou neho nazbierať. Pre normálnych užívateľov (zo strany zákazníkov) sú tieto zozbierané telemetrické dáta nepodstatné a vo svojej podstate nezmyselné, ale pre firmu Intelligent Technologies sú podsatným prínosom vo forme informácií pomocou ktorých môže ďalej vylepšovať svoje produkty pre zákazníkov.

Zoznam použitých zdrojov

- [1] O nás | Intelligent Technologies. *Intelligent Technologies / Business Intelligence v akcii* [online]. Brno: Intelligent Technologies, 2017 [cit. 2017-04-16]. Dostupné z: <http://www.intecs.cz/o-nas/>
- [2] TURLEY, Paul. *Professional Microsoft SQL server 2008 reporting services*. 3rd ed. Indianapolis, Ind.: Wiley Pub., c2009. Wrox professional guides. ISBN 978-0-470-24201-8.
- [3] KNIGHT, Brian. *Professional SQL server 2008 integration services*. Indianapolis, IN: Wiley Pub., c2008. ISBN 978-047-0247-952.
- [4] SIVAKUMAR HARINATH .. [ET AL.]. *Professional Microsoft SQL Server Analysis Services 2008 with MDX*. Indianapolis: Wiley, 2009. ISBN 978-047-0247-983.
- [5] SQL Server Integration Services | Microsoft Docs. *Microsoft* [online]. Redmond (Washington): Microsoft, 2017 [cit. 2017-05-08]. Dostupné z: <https://docs.microsoft.com/en-us/sql/integration-services/sql-server-integration-services>
- [6] DATABÁZY LS2006. *Univerzita komenského: Katedra aplikovanej matematiky a štatistiky* [online]. 2006 [cit. 2014-12-14]. Dostupné z: http://www.defm.fmph.uniba.sk/ludia/odrobina/Databazy_SQL_uvod.htm#_Toc381887624
- [7] Books Online. *Resources and Tools for IT Professionals / TechNet* [online]. Redmond (Washington): Microsoft, 2017 [cit. 2017-05-14]. Dostupné z: [https://technet.microsoft.com/en-us/library/ms130214\(v=sql.90\).aspx](https://technet.microsoft.com/en-us/library/ms130214(v=sql.90).aspx)

- [8] What is business intelligence? Resources and Tools for IT Professionals | TechNet [online]. Redmond (Washington): Microsoft, 2008 [cit. 2017-05-14]. Dostupné z: [https://technet.microsoft.com/en-us/library/cc811595\(v=office.12\).aspx](https://technet.microsoft.com/en-us/library/cc811595(v=office.12).aspx)
- [9] KIMBALL, Ralph a Margy ROSS. *The data warehouse toolkit: the definitive guide to dimensional modeling*. Third edition. ISBN 978-1-118-53080-1.

Zoznam obrázkov

| | |
|--|----|
| Obrázok č.1.: Logo spoločnosti, Zdroj: Intelligent technologies | 15 |
| Obrázok č.2.: MS SQL Server, Zdroj: technet.microsoft.com | 16 |
| Obrázok č.3.: MS SQL Database Engine, Zdroj: technet.microsoft.com..... | 17 |
| Obrázok č.4.: „This is not your father’s reporting tool.“ | 17 |
| Obrázok č.5.: MS SQL Server Reporting Services, Zdroj: technet.microsoft.com..... | 18 |
| Obrázok č.6.: MS SQL Server Analysis Services, Zdroj: technet.microsoft.com..... | 19 |
| Obrázok č.7.: MS SQL Server Integration Services, Zdroj: technet.microsoft.com | 20 |
| Obrázok č.8.: ETL proces, Zdroj: images.google.com | 22 |
| Obrázok č.9.: Dátová kocka, Zdroj: Vlastná tvorba | 27 |
| Obrázok č.10.: Olap Query Log Tabuľka, Zdroj: Vlastná tvorba | 40 |
| Obrázok č.11.: ETL pre využívanie reportov, Zdroj: Vlastná tvorba..... | 41 |
| Obrázok č.12.: Úloha toku dát pre načítanie reportov z katalógu, Zdroj: Vlastná tvorba | 42 |
| Obrázok č.13.: ETL pre zápis využívania kocky, Zdroj: Vlastná tvorba | 44 |

| | |
|--|----|
| Obrázok č.14.: ETL pre získanie užívateľov z Active Directory a ich zápis do tabuľky užívateľov, Zdroj: Vlastná tvorba | 46 |
| Obrázok č.15.: Načítanie užívateľov z AD, Zdroj: Vlastná tvorba | 46 |
| Obrázok č.16.: Úpdate pre tabuľku užívateľov, Zdroj: Vlastná tvorba | 47 |
| Obrázok č.17.: ETL pre naplnenie tabuliek dimenzií a tabuliek faktov | 48 |
| Obrázok č.18.: Data-flow task – detail plnenia faktovej tabuľky, Zdroj: Vlastná tvorba | 52 |
| Obrázok č.19.: Pohľad na dátový zdroj pri tvorbe multidimenzionálnej databázy, Zdroj: Vlastná tvorba | 53 |
| Obrázok č.20.: Multidimenzionálna databáza – kocka, Zdroj: Vlastná tvorba | 54 |
| Obrázok č.21.: Microsoft Visual Studio, Zroj: images.google.com | 55 |
| Obrázok č.22.: Power BI, Zdroj: images.google.com | 56 |
| Obrázok č.23.: Zobrazenie využívanie reportov v Power Bi, Zdroj: Vlastná tvorba | 56 |
| Obrázok č.24.: Zobrazenie využívanie reportov v Power Bi, Zdroj: Vlastná tvorba | 57 |
| Obrázok č.25.: Zobrazenie využívanie reportov v Power Bi, Zdroj: Vlastná tvorba | 58 |
| Obrázok č.26.: Zobrazenie využívanie reportov v Power Bi, Zdroj: Vlastná tvorba | 59 |

| | |
|---|----|
| Obrázok č.27.: Podiel formátov v akých boli reporty zobrazené, Zdroj: Vlastná tvorba | 60 |
| Obrázok č.28.: Zobrazenie užívateľov a reportov, ktoré si pozerali, Zdroj: Vlastná tvorba | 61 |
| Obrázok č.29.: Prostredie PowerBI, Zdroj: Vlastná tvorba | 62 |
| Obrázok č.30.: Report z prostredia Visual Studio, Zdroj: Vlastná tvorba..... | 63 |
| Obrázok č.31.: Report prehliadania OLAP servera, Zdroj: Vlastná tvorba..... | 64 |
| Obrázok č.32.: Prostredie Visual Studio, Zdroj: Vlastná tvorba | 66 |
| Obrázok č.33.: Prehliadanie OLAP kocky v exceli, Zdroj: Vlastná tvorba..... | 67 |

Poznámky